Sanna Hassinen
Katariina Hemmo
Timo Taskinen

LUKION LYHYT MATEMATIIKKA 5

SIGMA

Tilastot ja todennäköisyys

Sanoma Pro Oy • Helsinki

Tiedustelut
Sanoma Pro Oy, Helsinki
puh. 020 391 000
sähköposti: asiakaspalvelu@sanomapro.fi

www.sanomapro.fi

Kustantaja: Sanoma Pro Oy

Tämä aineisto on digitaalinen rinnakkaisversio tuotteesta
ISBN 978-952-63-1352-8 Sigma 5

12 kk lisenssi ISBN 978-952-63-2346-6
48 kk lisenssi ISBN 978-952-63-2348-0

Toimitus: Sanna Niemelä
Piirrokset: Marja Venäläinen ja Juho Niemelä
Ulkoasu: Riitta Piuva
Taitto: Eija Högman
Kansi: Riitta Piuva
Kannen kuva: Getty Images

Valokuvat: iStockphoto paitsi s. 20 Vastavalo

© Sanna Hassinen, Katariina Hemmo, Timo Taskinen ja Sanoma Pro Oy

Kopiointiehdot
Tämä teos on digitaalinen oppikirja, joka on suojattu tekijänoikeuslailla (404/61).

Sisällys

1 Tilastot 7

1.1 Tilastotieteen peruskäsitteitä 7
Tehtäviä 18
1.2 Summafrekvenssit 26
Tehtäviä 34
1.3 Keskiarvo 42
Tehtäviä 51
1.4 Keskihajonta 56
Tehtäviä 64
1.5 Normaalijakauma 68
Tehtäviä 83

2 Todennäköisyys 88

2.1 Malleja todennäköisyyden määrittämiseksi 88
Tehtäviä 99
2.2 Komplementti ja kertolaskusääntö 103
Tehtäviä 111
2.3 Yhteenlaskusääntö 115
Tehtäviä 122
2.4 Tuloperiaate ja kombinaatio 125
Tehtäviä 132
2.5 Binomitodennäköisyys 136
Tehtäviä 141

3 Kertausosa 145

4 Ekstrat 157

Tehtävien vastaukset 188

Hakemisto 197

Kirjan käyttäjälle

Matematiikan opetuksen tarkoituksena on rohkaista opiskelijaa oivaltavaan oppimiseen, kehittää matemaattista ajattelua ja ongelmanratkaisutaitoa. Sigman rakenne tukee näiden tavoitteiden toteutumista.

Pohdinta

Osa kirjan alaluvuista alkaa pohdintatehtävillä. Niiden tarkoituksena on yleensä johdatella opetettavaan aiheeseen ja herättää kysymyksiä. Opettaja voi käyttää tehtäviä osana opetusta, tai opiskelija voi tutustua tehtäviin itsenäisesti.

Teoria

Teoria on jaettu aihekokonaisuuksiin sivuotsikoiden avulla. Näin eri osa-alueet löytyvät helpommin. Suuremmat kokonaisuudet voidaan otsikoiden avulla jakaa sopiviin opetuskokonaisuuksiin.

Tehtävät

Tehtävät on koottu teorian mukaisten sivuotsikoiden avulla kokonaisuuksiksi, joista on helppo valikoida sopivat tehtävät aihepiirin ja vaikeusasteen mukaan.

120. Vaativammat tehtävät on merkitty väripohjalla.

E1 Joidenkin harjoitustehtävien edessä olevan esimerkkiviittauksen tarkoituksena on tukea harjoittelemista sekä oppitunnilla että kotona. Jokaista kirjan esimerkkiä vastaa tehtäväosassa samantyyppinen tehtävä.

Kirjassa on mukana vanhoja ylioppilastehtäviä. Ylioppilaskokeen vuosi ja tehtävänumero on ilmoitettu tehtävänannon lopussa, esimerkiksi [S2008, 1]

Sekalaisia tehtäviä

Tehtäväosion lopussa on sekalaisia tehtäviä -osio, johon on koottu muutama tehtävä koko luvun aihealueista. Tämän osion tehtäviä voi käyttää esimerkiksi kotitehtävinä.

Keskeisiä käsitteitä

Jokaisen luvun lopussa on käsitekartta, joka kokoaa luvun keskeiset käsitteet.

Kertausosa

Kirjan lopussa olevaan kertausosaan on koottu lyhyesti kunkin luvun tärkeimmät asiat. Jokaiseen lukuun liittyy lisäksi kertaustehtäviä. Kertausosan lopussa on myös kolme harjoituskoetta.

Ekstrat

Kirjan loppuun on koottu ekstraluvuksi kurssiin liittyvää oheismateriaalia. Ekstroja voi tarpeen mukaan käyttää joko opetuksessa tai opiskelijan omana lukemistona. Teoriaosassa on viittauksia ekstralukuun.

Helsingissä kesäkuussa 2013

Tekijät

Projektityö: Tilastollinen tutkimus

Tilasto-osuudesta on mahdollista tehdä projektityö, jonka tarkoituksena on

- syventää tilastotieteen peruskäsitteiden ymmärtämistä
- harjoitella tilastollisten tunnuslukujen laskemista
- pohtia tilastollisen tutkimuksen tekemiseen liittyviä ongelmia
- opettaa aineiston analysointia tilastollisten tunnuslukujen avulla.

Projektityö: Projektityön ohjeita s. 157

Projektityössä kerätään itse tutkittava aineisto ja analysoidaan sitä frekvenssijakaumien sekä erilaisten tunnuslukujen avulla. Jotta projektityö ehtii valmistumaan kurssin aikana, sen tekeminen kannattaa aloittaa heti kurssin alussa. Tutkimusta voi täydentää aina sitä mukaa, kun kurssilla käsitellään uusia asioita. Projektityön tekemiseen liittyvät ohjeet löytyvät kirjan takaa Ekstraosasta.

Ennen tutkimuksen tekemistä kannattaa pohtia yhdessä opettajan kanssa:

Tehdäänkö tutkimus yksin vai parityönä?

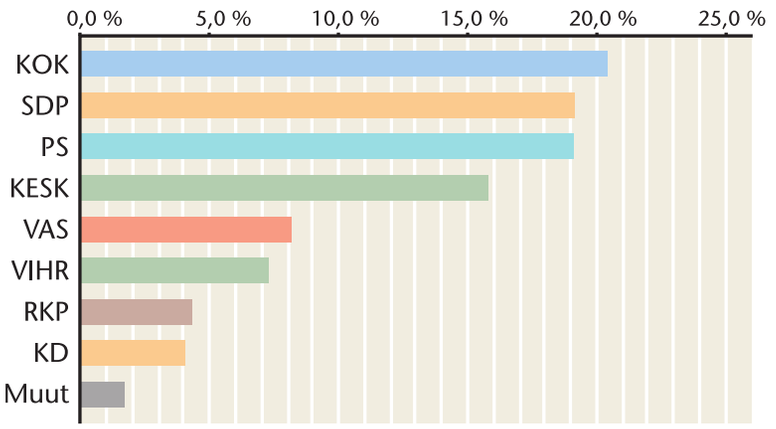
- Käytetäänkö tutkimuksen tekemiseen tietokonetta?
- Kuinka paljon tutkimusta tehdään tunnilla?
- Onko tutkimus pakollinen vai vapaaehtoinen?
- Mikä on tutkimuksen aikataulu?
- Tehdäänkö tutkimuksesta esimerkiksi raportti tai juliste?
- Miten tutkimus arvostellaan?
- Miten tutkimus vaikuttaa kurssiarvosanaan?
- Miten tutkimuksesta annetaan palautetta?

1 Tilastot

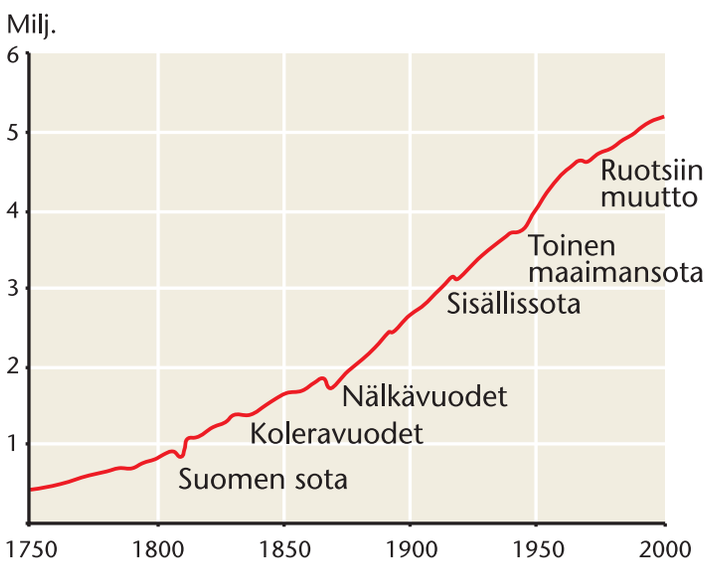
1.1 Tilastotieteen peruskäsitteitä

Tilastollinen havaintoaineisto

Sanasta **tilastotiede** tulee useimmille mieleen erilaiset taulukot ja kuvaajat, kuten väestönkasvu vuosittain tai puolueiden kannatusosuudet.



Puolueiden suhteelliset äänimäärät eduskuntavaaleissa 2011
(Lähde: Tilastokeskus)



Suomen väestön kehitys 1750–2000 (Lähde: Tilastokeskus)

Kuvaajat ovat kuitenkin vain pieni osa tilastotiedettä. Tilastotieteessä käsitellään menetelmiä, joita tarvitaan kokemusperäisen, numeerisen tietoaineiston hankinnassa, kuvaamisessa ja analyysissä sekä saatujen tulosten tulkinnassa.

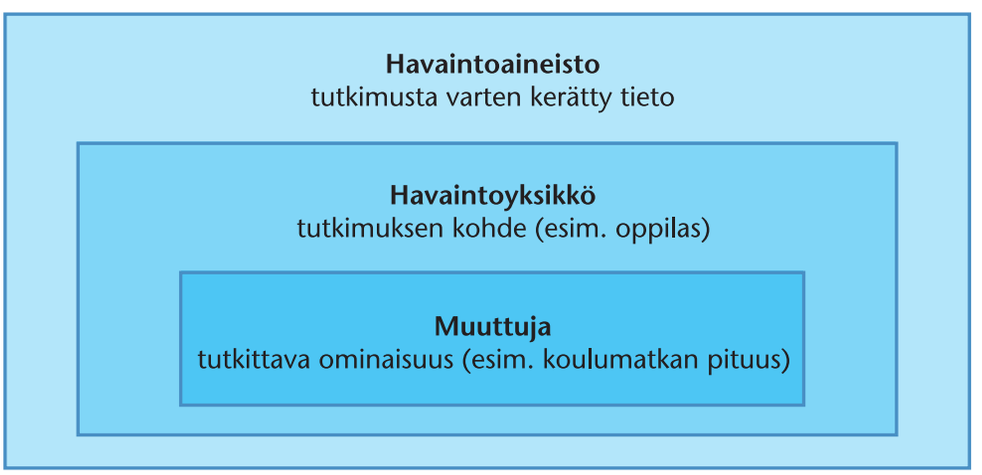
Tilastotieteen avulla voidaan hallita numeerista tietoa, käsitellä epävarmuuksia, yleistää erilaisia tuloksia, analysoida riippuvuuksia ja ennustaa tulevia tilanteita. Tilastotieteen keinojen avulla halutaan useimmiten muodostaa mahdollisimman todenmukainen kuva tarkasteltavasta tilanteesta päätöksenteon tueksi.

Havaintoaineiston kerääminen s. 160

Tilastotieteessä tehtävä tutkimus on ensin suunniteltava hyvin. On mietittävä, mitä asioita halutaan tutkia ja miten tutkittavaa tietoa kerätään. Tutkimusta varten kerättyä tietoa kutsutaan **havaintoaineistoksi** tai **tilastoaineistoksi**.

Muuttujien luokittelu s. 164

Havaintoaineisto koostuu **havainto**- eli **tilastoyksiköistä**. Jokaiseen havaintoyksikköön liittyy tutkittavia ominaisuuksia eli **muuttujia**. Jos tutkitaan esimerkiksi opetusryhmän oppilaiden koulumatkojen pituuksia, oppilaat ovat havaintoyksikköjä ja koulumatkan pituus on muuttuja. Nämä tiedot yhdessä muodostavat havaintoaineiston.



Perusjoukko eli **populaatio** koostuu kaikista havaintoyksiköistä, joista tietoa halutaan kerätä. Jos esimerkiksi tutkitaan tietyn opetusryhmän oppilaiden koulumatkan pituutta, perusjoukkona ovat kaikki ryhmän oppilaat.

Jos havaintoaineistoon kuuluvat kaikki perusjoukon havaintoyksiköt, sitä kutsutaan **kokonaisaineistoksi**. Oppilaiden koulumatkan pituuden tutkimisessa tämä tarkoittaa, että tarvittavat tiedot kerätään kaikilta ryhmän oppilailta.

Kokonaisaineistoa on usein mahdotonta kerätä, eikä se useimmiten ole tarpeenkaan riittävän tarkkojen johtopäätösten tekemiseksi. Tästä syystä kokonaisaineiston sijasta tutkitaan usein **otosta**. Otos on perusjoukon osa, joka on kerätty jollain **otantamenetelmällä**. Otantamenetelmät ovat erilaisia tapoja poimia havaintoyksiköitä perusjoukosta niin, että otos kuvaisi perusjoukkoa mahdollisimman hyvin ja puolueettomasti.

Erilaisia frekvenssejä

Havaintoaineiston keräämisen jälkeen aineistoa käsitellään niin, että kaikki olennainen tieto saadaan näkyviin. Havaintoaineiston käsittely aloitetaan yleensä frekvenssien laskemisella. **Frekvenssi** f kertoo, kuinka monta kertaa muuttujan arvo esiintyy havaintoaineistossa.

Seuraavassa taulukossa ovat erään oppilasryhmän sisarusten lukumäärät.

0	3	2	2	1	1	1	1	1	2	1
2	1	0	1	2	1	0	2	1	1	3

Frekvenssin määrittämiseksi lasketaan muuttujan arvojen esiintymiskerrat. Apuna voi käyttää esimerkiksi tukkimiehen kirjanpitoa.

Sisarusten lukumäärä		Frekvenssi f
0	𝍸	3
1	𝍸 𝍸 |	11
2	𝍸 |	6
3	||	2

Diskreetti eli epäjatkuva muuttuja voi saada vain tiettyjä arvoja, esim. kokonaislukuja.

Tällä tavalla laadittu taulukko, jossa esiintyvät muuttujien arvot ja niihin liittyvät frekvenssit, on nimeltään **frekvenssijakauma**. Lukumäärää kuvaavasta frekvenssistä voidaan käyttää myös nimitystä **absoluuttinen frekvenssi**. Se kertoo esiintymiskertojen lukumäärän, kun taas **suhteellinen frekvenssi** f % kertoo esiintymiskertojen prosentuaalisen osuuden havaintoyksiköiden kokonaismäärästä.

Frekvenssit

- Frekvenssi f on muuttujan arvojen esiintymiskertojen lukumäärä.
- Suhteellinen frekvenssi f % on frekvenssin prosentuaalinen osuus havaintoyksiköiden lukumäärästä.
- Frekvenssijakauma nähdään taulukosta, jossa ovat muuttujan arvot ja niihin liittyvät frekvenssit.

Sisarusten lukumäärä	Frekvenssi f	Suhteellinen frekvenssi f %
0	3	$\frac{3}{22} = 0{,}136\ldots \approx 14\ \%$
1	11	$\frac{11}{22} = 0{,}5 = 50\ \%$
2	6	$\frac{6}{22} = 0{,}2727\ldots \approx 27\ \%$
3	2	$\frac{2}{22} = 0{,}0909\ldots \approx 9\ \%$

Havaintoyksiköiden lukumäärä
$3 + 11 + 6 + 2 = 22$

Suhteellisten frekvenssien avulla on helpompi vertailla aineistoja, joissa havaintoyksiköitä on eri määrät.

Frekvenssijakaumasta nähdään, että muuttujan arvolla 1 on suurin frekvenssi eli 11. Tätä muuttujan arvoa kutsutaan aineiston **tyyppiarvoksi** eli **moodiksi** Mo. Moodi on esimerkki tilastollisista tunnusluvuista. Nämä luvut kuvaavat muuttujien jakaumien keskeisiä ominaisuuksia.

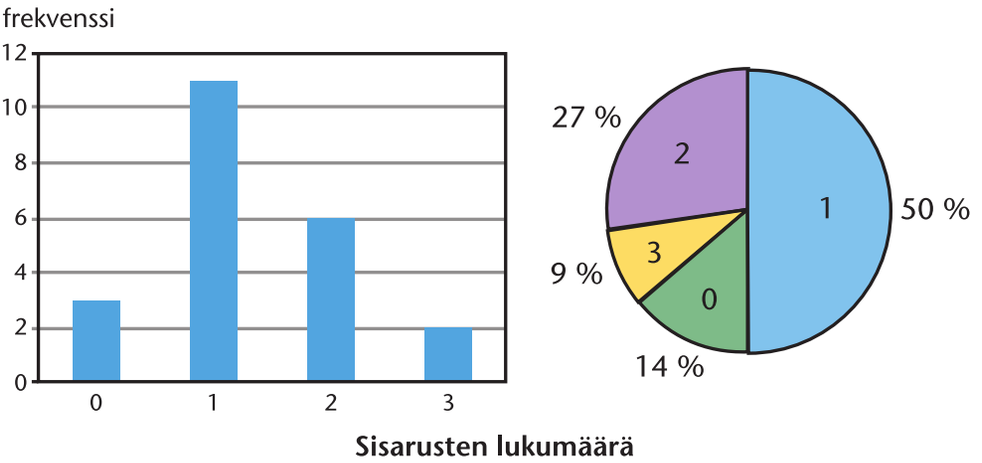
Moodi Mo

Muuttujan arvo, jolla on suurin frekvenssi.

Frekvenssien kuvaaminen s. 168

Tilastollista aineistoa voidaan havainnollistaa erilaisten kuvaajien avulla. Näistä tavallisimmat ovat pylväsdiagrammi ja sektoridiagrammi.

Diskreetin muuttujan pylväät piirretään erilleen.



Jos pylväsdiagrammissa pylväät ovat keskenään yhtä leveitä, pylvään korkeus määräytyy frekvenssin tai suhteellisen frekvenssin mukaan. Sektoridiagrammilla kuvataan yleensä suhteellisia frekvenssejä. Sektorin keskuskulman suuruus määräytyy suhteellisen frekvenssin mukaan.

ESIMERKKI 1

Talitiainen (*Parus Major*) on Euraasian yleisimpiä lintulajeja. Se munii 6–12 munaa. Luonnonpuistossa tutkittiin talitiaispoikueiden kokoja laskemalla munat 20 pesästä. Tulokset olivat seuraavat:

9	6	9	8	11	9	8	7	9	10
12	11	11	6	7	7	11	9	8	9

a) Laadi aineistosta frekvenssijakaumat ja määritä aineiston moodi.
b) Havainnollista aineistoa sektoridiagrammilla.

RATKAISU

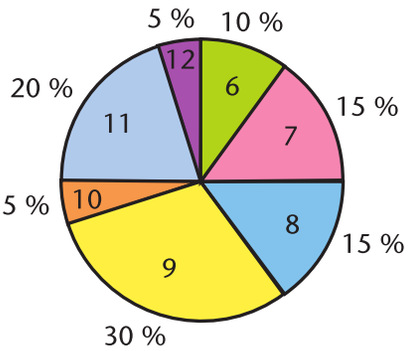
a) Kootaan aineisto taulukkoon ja lasketaan suhteelliset frekvenssit.

Munia (kpl)	*f*	*f* %
6	2	$\frac{2}{20} = 0{,}1 = 10\ \%$
7	3	$\frac{3}{20} = 0{,}15 = 15\ \%$
8	3	$\frac{3}{20} = 0{,}15 = 15\ \%$
9	6	$\frac{6}{20} = 0{,}3 = 30\ \%$
10	1	$\frac{1}{20} = 0{,}05 = 5\ \%$
11	4	$\frac{4}{20} = 0{,}2 = 20\ \%$
12	1	$\frac{1}{20} = 0{,}05 = 5\ \%$

Koska muuttujan arvolla 9 on suurin frekvenssi, Mo = 9.

b) Sektoridiagrammin piirtämiseksi on laskettava jokaista muuttujan arvoa vastaavan sektorin keskuskulma. Koko ympyrän asteluku on 360°.

Munia (kpl)	*f* %	Keskuskulma
6	10 %	0,1 · 360° = 36°
7	15 %	0,15 · 360° = 54°
8	15 %	0,15 · 360° = 54°
9	30 %	0,3 · 360° = 108°
10	5 %	0,05 · 360° = 18°
11	20 %	0,2 · 360° = 72°
12	5 %	0,05 · 360° = 18°



VASTAUS a) Mo = 9

Aineiston luokittelu

Jatkuva muuttuja, esimerkiksi pituus, saa useita eri arvoja. Tällainen aineisto kannattaa yleensä **luokitella**, koska luokittelun avulla aineiston ominaisuudet saadaan paremmin esiin. Luokittelussa tietyn suuruiset muuttujan arvot kootaan yhteen.

POHDINTA 1

Aapo tutki kotipihan omenapuun omenasatoa. Hän keräsi puun alta satunnaisesti 20 omenaa ja punnitsi ne. Omenoiden massat olivat seuraavat:

110 g	91 g	112 g	89 g	90 g
84 g	72 g	95 g	108 g	95 g
102 g	89 g	107 g	88 g	115 g
112 g	118 g	125 g	85 g	77 g

Täytä seuraava taulukko esimerkin mukaan ja vastaa kysymyksiin.

Massa (g)	Kuinka monta tämän painoista omenaa on otoksessa?	Kuinka monta prosenttia omenamäärä on koko otoksen koosta?
70–79	2	$\frac{2}{20} = 0{,}1 = 10\ \%$
80–89		
90–99		
100–109		
110–119		
120–129		

Kuinka monta omenaa painoi
a) alle 100 grammaa b) alle 120 grammaa?

Kuinka monta prosenttia omenista painoi
a) alle 90 grammaa b) alle 120 grammaa?

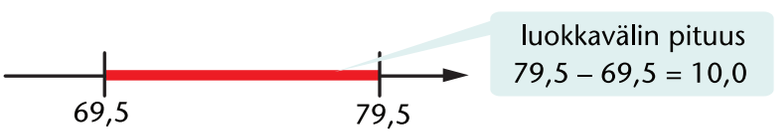
Luokkia muodostetaan yleensä 4–10 kappaletta havaintoaineiston laajuudesta riippuen.

Luokka ilmoitetaan yleensä **luokkavälinä** eli ilmoitetaan luokan **alaraja** ja **yläraja**. Useimmiten taulukossa ilmoitetut luokkarajat eivät ole **todellisia luokkarajoja**. Todellisissa luokkarajoissa otetaan huomioon esimerkiksi mittaustarkkuus ja mahdolliset pyöristykset.

Pohdintatehtävässä ensimmäiseen luokkaan kuuluvat omenat, joiden massa on 70 g – 79 g. Omenien massat on tässä pyöristetty gramman tarkkuuteen. Jos omenan massa on vähintään 69,5 g, se kuuluu tähän luokkaan. Myös ne omenat, joiden massa on alle 79,5 g, kuuluvat tähän luokkaan. Nämä pyöristysrajat ovat luokan todelliset luokkarajat.

Massa (g)	Todellinen alaraja (g)	Todellinen yläraja (g)
70–79	69,5	79,5
80–89	79,5	89,5

Luokkavälin pituus on todellisten luokkarajojen erotus.



Luokat kannattaa valita tasavälisiksi niin, että luokkavälin pituus on aina sama. Jokainen muuttujan arvo saa kuulua vain yhteen luokkaan.

Luokitellussa aineistossa tarvitaan usein yksi luku edustamaan kutakin luokkaa. Se on yleensä luokan keskellä oleva luku eli **luokkakeskus**. Luokkakeskus on luokan todellisen alarajan ja todellisen ylärajan keskiarvo. Luokkakeskusta tarvitaan esimerkiksi erilaisten tunnuslukujen laskemiseen ja luokitellun aineiston kuvaajien tekemiseen. Esimerkiksi luokitellun aineiston moodina on sen luokan luokkakeskus, jolla on suurin frekvenssi.

Luokkakeskuksen laskeminen

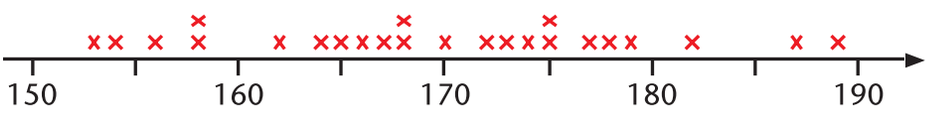
$$\text{luokkakeskus} = \frac{\text{todellinen alaraja} + \text{todellinen yläraja}}{2}$$

ESIMERKKI 2 Taulukossa on erään oppilasryhmän oppilaiden pituudet senttimetreinä:

154	178	162	156	168	189	165	166
168	174	175	182	187	153	167	172
173	158	170	179	175	164	158	177

Luokittele aineisto neljään luokkaan ja määritä jokaisen luokan luokkakeskus.

RATKAISU Luokittelun helpottamiseksi kannattaa piirtää pistekuvio eli sijoittaa pituudet lukusuoralle. Siitä nähdään, mille välille pituudet asettuvat ja voidaan valita sopiva pituus luokalle.



Kaikki muuttujan arvot sijoittuvat välille 150–190. Tämän välin pituus on 40 cm, joten muodostetaan luokkia, joissa luokkavälin pituus on 10 cm.

Peräkkäisten luokkien ylä- ja alarajat ovat eri suuret.

Alaraja	Yläraja	Todellinen alaraja	Todellinen yläraja	Luokkakeskus	Luokka
151	160	150,5	160,5	$\frac{150{,}5+160{,}5}{2}=155{,}5$	151–160
161	170	160,5	170,5	$\frac{160{,}5+170{,}5}{2}=165{,}5$	161–170
171	180	170,5	180,5	$\frac{170{,}5+180{,}5}{2}=175{,}5$	171–180
181	190	180,5	190,5	$\frac{180{,}5+190{,}5}{2}=185{,}5$	181–190

ESIMERKKI 3

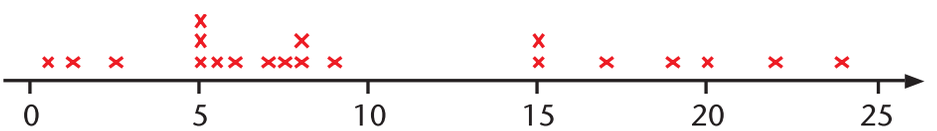
Matematiikan kurssin oppilaiden koulumatkojen pituudet (km) olivat seuraavat:

0,5	5	7,5	15	8	8	5	9	20	24
15	7	17	5	1,2	22	2,5	19	5,5	6

Luokittele aineisto viiteen luokkaan siten, että ensimmäinen luokka on 0–4 km ja määritä luokkakeskukset. Muodosta frekvenssijakaumat ja havainnollista suhteellista frekvenssijakaumaa kuvaajan avulla.

RATKAISU

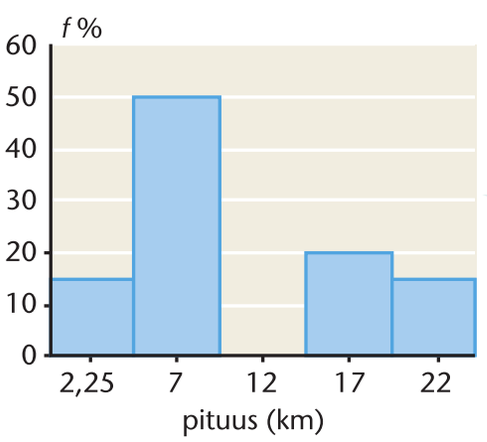
Luokittelun helpottamiseksi piirretään pistekuvio, josta nähdään, moneenko luokkaan aineisto kannattaa jakaa.



Luokka (km)	Luokkakeskus (km)	f	f %
0–4	$\frac{0+4,5}{2} = 2,25$	3	15 %
5–9	$\frac{4,5+9,5}{2} = 7$	10	50 %
10–14	$\frac{9,5+14,5}{2} = 12$	0	0 %
15–19	$\frac{14,5+19,5}{2} = 17$	4	20 %
20–24	$\frac{19,5+24,5}{2} = 22$	3	15 %

Pylvään korkeus määräytyy frekvenssin tai suhteellisen frekvenssin mukaan.

Piirretään pylväät keskenään yhtä leveiksi. Koska pituus on jatkuva muuttuja, pylväät piirretään kuvaajassa yhteen. Merkitään pylvään keskikohdalle luokkakeskus.



Jatkuvan muuttujan pylväsdiagrammia kutsutaan histogrammiksi.

Tilastot ja todennäköisyys

Tilastoja käytetään erilaisten ennusteiden laatimiseen. Monissa ilmiöissä tapahtuman suhteellinen frekvenssi kertoo tapahtuman **todennäköisyyden**. Esimerkki tällaisesta **satunnaisilmiöstä** on kolikon heittäminen. Todennäköisyyttä, että kolikkoa heitettäessä saadaan kruuna, arvioidaan heittämällä kolikkoa useita kertoja ja taulukoimalla lopputulokset.

Englantilainen tilastotieteilijä *Karl Pearson* (1857–1936) heitti lanttia 12 000 kertaa ja sai 6019 kruunaa. Kruunan prosentuaalinen osuus oli $\frac{6019}{12\,000} \approx 50{,}2\,\%$.

Voidaankin sanoa, että heitettäessä kolikkoa kruunan todennäköisyys on likimain 50 % = 0,5. Tämä merkitään matematiikassa symbolin P avulla seuraavasti:

$$P(\text{kolikonheitossa saadaan kruuna}) = 0{,}5$$

Kun todennäköisyyden laskeminen perustuu tilastoon, puhutaan **tilastollisesta todennäköisyydestä**.

Esimerkki 4

Väärennettyä noppaa heitetään 500 kertaa. Silmälukujen esiintymiskerrat on merkitty taulukkoon:

Silmäluku	1	2	3	4	5	6
f	43	40	45	85	120	167

Laske todennäköisyys, että noppaa heitettäessä saadaan silmäluvuksi a) ykkönen b) kuutonen.

Ratkaisu

a) Lasketaan silmäluvun 1 suhteellinen frekvenssi:

$$\frac{43}{500} \cdot 100\,\% = 8{,}6\,\%$$

$$P(\text{saadaan ykkönen}) = 0{,}086$$

b) Lasketaan silmäluvun 6 suhteellinen frekvenssi:

$$\frac{167}{500} \cdot 100\,\% = 33{,}4\,\%$$

$$P(\text{saadaan kuutonen}) = 0{,}334$$

Vastaus

a) 0,086 b) 0,334

Tehtäviä

Erilaisia frekvenssejä

1. Taulukossa on Suomen kuuden laajalevikkisimmän sanomalehden levikit vuonna 2011. Laske suhteelliset frekvenssit.

Lehti	Levikki
Helsingin Sanomat, Helsinki	365 994
Ilta-Sanomat, Helsinki	143 321
Aamulehti, Tampere	130 081
Turun Sanomat, Turku	103 314
Iltalehti, Helsinki	102 124
Maaseudun tulevaisuus	83 259
Muut sanomalehdet yhteensä	1 767 109

Lähde: Levikintarkastus Oy

a) Kuinka suuri prosenttiosuus on Helsingin Sanomilla?
b) Kuinka suuri kappalemääräinen levikki on valtakunnan kolmella suurimmalla lehdellä?
c) Kuinka suuri prosentuaalinen osuus on valtakunnan neljällä suurimmalla lehdellä?

2. Taulukossa kerrotaan, miten opiskelijat jakaantuivat eri oppilaitoksiin vuonna 2001. Muodosta frekvenssijakauma ja suhteellinen frekvenssijakauma.
Kuinka monta prosenttia opiskelijoista opiskelee
a) peruskoulussa tai lukiossa b) muualla kuin yliopistossa?

Oppilaitostyyppi	Opiskelijoita
peruskoulut	595 727
lukiot	128 642
ammatillinen koulutus	160 115
ammattikorkeakoulutus	118 013
yliopistokoulutus	162 939
yhteensä	1 165 436

Lähde: Tilastokeskus

E1 3. Kemian toiselle kurssille osallistui 16 opiskelijaa. Heidän arvosanansa tältä kurssilta on esitetty taulukossa:

6	10	8	8	6	7	8	9
7	9	6	7	7	9	7	10

a) Muodosta aineistosta frekvenssijakaumat ja määritä aineiston moodi.
b) Havainnollista aineistoa sektoridiagrammilla.

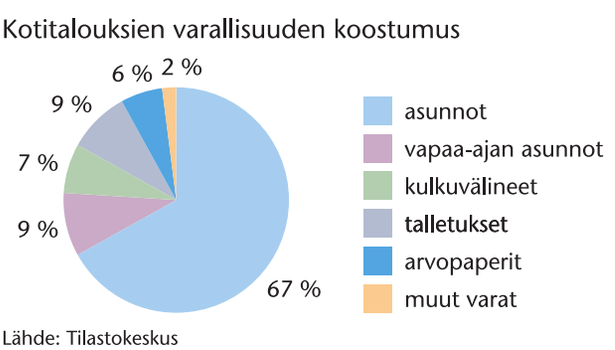
4. Suomalaisten sijoitukset jääkiekon MM-kilpailuissa vuosina 1939–2012 olivat seuraavat:

Sijoitus	1.	2.	3.	4.	5.	6.	7.	9.	13.
Kpl	2	6	3	11	16	8	11	1	1

Lähde: Suomen Jääkiekkomuseo

a) Muodosta frekvenssijakaumat.
b) Havainnollista aineistoa pylväsdiagrammilla.

5. Nelihenkisen perheen nettovarallisuudeksi arvioitiin 248 000 euroa. Laske viereisen sektoridiagrammin avulla perheen
a) asunnon arvo
b) talletusten määrä
c) arvopaperisijoitusten määrä.



Lähde: Tilastokeskus

6. Työpaikalla tehtiin henkilöstön kuntokartoitusta. Työntekijöiltä kysyttiin, kuinka monena päivänä viikossa kukin harrastaa liikuntaa vähintään tunnin ajan kerrallaan. Kyselyyn vastasi 20 työntekijää.

2	1	2	2	4	0	2	3	3	5
2	1	0	5	6	5	3	2	3	1

a) Laadi frekvenssijakaumat.
b) Määritä aineiston moodi.
c) Piirrä aineistosta sektoridiagrammi.

Aineiston luokittelu

7. Laske luokkien luokkakeskukset ja luokkavälin pituus. Mitkä ovat todelliset luokkarajat?

a)

Paino (kg)
50–54
55–59
60–64

b)

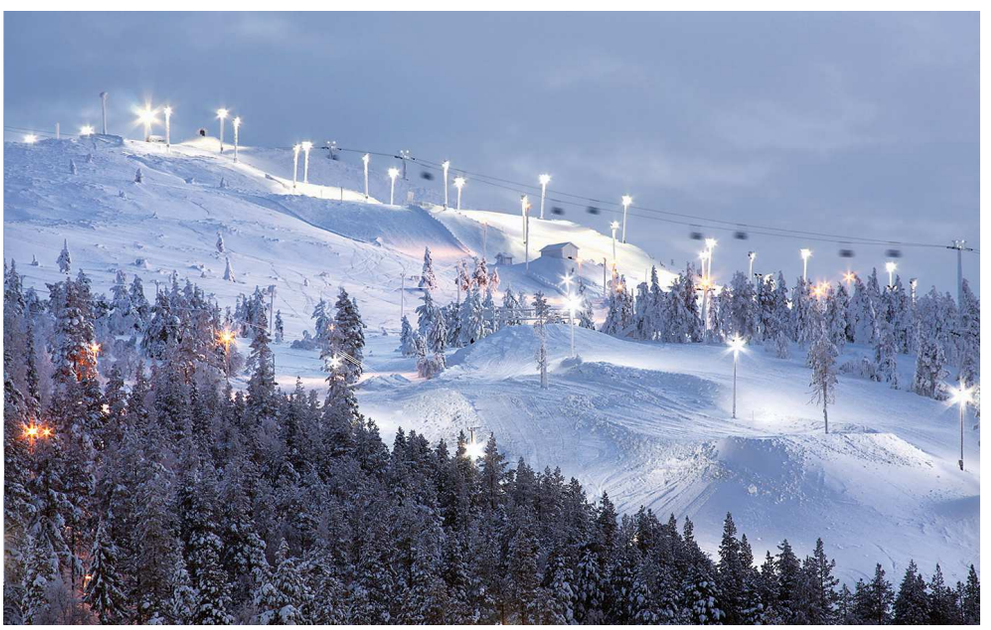
5 km juoksuaika (min)
24–29
30–35
36–41

E2 8. Luokittele Levin hiihtokeskusten hissien korkeuseroja kuvaava aineisto viiteen luokkaan.

a) Mitkä ovat todelliset luokkarajat?
b) Laske luokkien luokkakeskukset.
c) Laske luokkavälin pituus.

Hissin nimi	Korkeusero (m)
Koutalaki	154
Hemmo	128
Marjalaki	159
Rakka	66
Jäkälälaki	245
Taalo	290
Utsuvaara	160
Sammu	171

Hissin nimi	Korkeusero (m)
Etelärinne	233
Huippu	133
Tieva	140
Himmeriikki	152
Lammaskuru	188
Vesirikka-autto	132
Tunneli	162
Gondoli	302



9. Luokittele MM-rallin osakilpailuvoittajien keskinopeuksiin (2001) liittyvä aineisto kuuteen luokkaan ja laske luokkien luokkakeskukset.

Osakilpailu	Keskinopeus (km/h)
Suomi	119,9
Kenia	119,3
Ruotsi	110,1
Australia	107,0
Britannia	104,4
Uusi-Seelanti	100,8
Katalonia	94,3
Korsika	92,1
Argentiina	90,4
San Remo	89,8
Kreikka	82,4
Portugali	79,6
Monte Carlo	76,7
Kypros	66,6

Lähde: World Rallying

10. Jalkapalloseura Valkeakosken Hakan SM-sarjasijoitukset vuosina 1980–2000 olivat seuraavat:

2. 3. 3. 3. 4. 8. 5. 9. 6. 4.
8. 3. 6. 10. 6. 1. 11. 1. 1. 1.

Luokittele aineisto neljään luokkaan ja muodosta frekvenssijakauma.
a) Laske luokkien luokkakeskukset.
b) Mitkä ovat luokkien todelliset luokkarajat?

E3 11. Alla olevaan taulukkoon on koottu Etelä-Amerikan valtioiden eliniän odotteita. Luokittele aineisto siten, että alimpana luokkana on 65–67. Määritä luokkien luokkakeskukset. Muodosta frekvenssijakaumat ja havainnollista aineistoa kuvaajan avulla.

77,6	77,3	77,3	76,6	76,3	75,8	75,3	74,1	73,7	73,6
72,3	72	71,5	70,7	70,5	70,3	68,2	66,9	66,3	

12. Taulukossa on esitetty 30 Euroopan valtion inflaatioprosentit vuonna 2012. Luokittele aineisto siten, että ensimmäisenä luokkana on 0,1–1,0. Määritä luokkien luokkakeskukset ja aineiston moodi.

0,4	0,9	1,0	1,9	2,1	2,2	2,3	2,4	2,4	2,4
2,6	2,6	2,8	2,8	2,8	2,8	2,9	3,1	3,2	3,2
3,2	3,2	3,3	3,4	3,5	3,7	3,7	4,2	5,7	6,0

Lähde: Eurostat

13. Vaasan vaalipiirin kuntien äänestysprosentit eduskuntavaaleissa 2012 on esitetty alla olevassa taulukossa:

58,1	58,2	58,4	58,9	59,3	60,2	61,4	63	63,2	63,5
63,5	64,1	64,1	64,9	65	65,8	66,4	66,7	66,7	67,3
67,3	67,7	68	68,3	68,4	68,6	68,8	69	69,1	69,5
69,8	71	71,4	71,9	72,4	73,6	74,8	75,3	77,9	79,9

Lähde: Oikeusministeriö

Täydennä aineiston avulla alla oleva taulukko.

Äänestysprosentti	*f*	*f* %	Luokkakeskus
56–60			
61–65			
66–70			
71–75			
76–80			

Tilastot ja todennäköisyys

14. Raha-arpajaisissa lasketaan liikkeelle 20 000 arpaa. Voitot jakautuvat seuraavan taulukon mukaisesti:

Voitto (€)	0	10	50	500	5000
Kpl	12 000	6000	1000	900	100

Henkilö ostaa yhden arvan. Laske aineiston avulla todennäköisyys, että hän
a) ei saa lainkaan voittoa
b) voittaa päävoiton.

15. Laboratoriossa tutkittiin risteytyskokeiden avulla hiirten turkin väriä. Väreissä esiintyivät seuraavat muunnokset:

Värimuunnos	Hiirten määrä
valkoinen	255
musta	87
ruskea	102
monivärinen	62

Määritä aineiston avulla todennäköisyys, että satunnaisesti valittu hiiri on
a) monivärinen
b) yksivärinen.

16. Länsi-Suomessa sijaitsevalla sääasemalla tutkittiin viiden vuoden ajan heinäkuun sadepäivien esiintymistiheyttä ja saatiin seuraavat tulokset:

Vuosi	Sadepäiviä (kpl)
2007	9
2008	7
2009	7
2010	10
2011	8

Laske aineiston perusteella todennäköisyys, että mainitulla alueella satunnaisesti valittuna heinäkuun päivänä sataa vettä.

Sekalaisia tehtäviä

17. Erään golfradan vaikein reikä sijaitsee hiekkakentän takana. Radalla pidettyyn kilpailuun osallistui 16 pelaajaa. Heidän lyöntiensä määrät tällä rei'ällä on esitetty alla olevassa taulukossa:

5	4	5	5	7	5	7	9
7	8	8	6	5	6	7	9

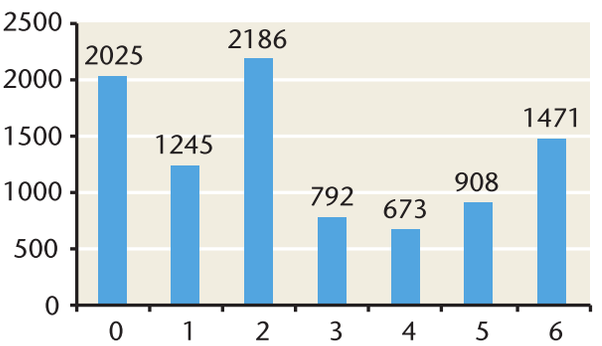
a) Laadi aineistosta frekvenssijakaumat ja määritä aineiston moodi.
b) Havainnollista aineistoa sektoridiagrammilla.

18. Leijonalaumassa on yleensä 5–20 leijonaa. Kenialaisessa kansallispuistossa kartoitettiin alueella elävien leijonalaumojen kokoja. Lasketussa 22 laumassa oli leijonia seuraavasti:

17	11	12	7	15	8	19	8	5	10	12
12	16	7	6	9	9	14	13	9	8	18

Luokittele aineisto ja laadi frekvenssijakaumat. Määritä aineiston moodi.

19. Alla oleva pylväsdiagrammi esittää kevään 2007 lyhyen matematiikan ylioppilaskokeen tehtävän 8 pistejakaumaa (vaaka-akselilla pisteet, pylväiden korkeus osoittaa kyseisen pistemäärän saaneiden lukumäärän). Laadi vastaava sektoridiagrammi, josta ilmenee kunkin pistemäärän saaneiden suhteellinen osuus, ja ilmoita osuudet prosentteina. [K2009, 4]



20. Sairaalan synnytysosastolla koottiin viikon aikana syntyneiden poikavauvojen syntymäpainot (kg):

3,12	4,46	3,50	3,44	3,86
4,05	4,02	2,96	2,92	3,24
3,60	3,71	3,02	2,76	4,66

a) Luokittele aineisto viiteen luokkaan siten, että alimpana luokkana on 2,6–3,0. Laadi frekvenssijakaumat ja määritä luokkakeskukset.
b) Havainnollista aineistoa histogrammilla.

21. Suomen Punaisen Ristin veripalvelun tilastojen mukaan suomalaisten veriryhmäjakauma on seuraava:

Veriryhmä	Rhesus + (milj. henkilöä)	Rhesus – (milj. henkilöä)
A	2,01	0,27
B	0,81	0,11
O	1,52	0,27
AB	0,38	0,05

Millä todennäköisyydellä satunnaisesti valitun suomalaisen henkilön veriryhmä on
a) AB
b) Rhesus-positiivinen?

1.2 Summafrekvenssit

POHDINTA 1

Asukasyhdistys teki tutkimuksen alueellaan asuvien nuorten harrastuskertojen määristä viikossa. Tutkimustulokset on koottu seuraavaan taulukkoon.

Täydennä taulukko mallin mukaan.

Harrastus-kertoja/viikko	Frekvenssi *f*	Summa-frekvenssi *sf*	Suhteellinen summafrekvenssi *sf* %
0	12	12	$\frac{12}{138} = 0{,}0869\ldots \approx 9\ \%$
1	35	$12 + 35 = 47$	$\frac{47}{138} = 0{,}3405\ldots \approx 34\ \%$
2	18	$47 + 18 = 65$	$\frac{65}{138} = 0{,}4710\ldots \approx 47\ \%$
3	45		
4	16		
5	4		
6	8		

Vastaa taulukon perusteella seuraaviin kysymyksiin:

a) Kuinka moni nuori käy harrastuksissa kolme kertaa viikossa?

b) Kuinka monella prosentilla on korkeintaan kaksi harrastus-kertaa viikon aikana?

c) Kuinka monella nuorella on korkeintaan kolme harrastusta?

d) Kuinka monella nuorella on vähintään neljä harrastuskertaa?

e) Kuinka monta prosenttia nuorista käy vähintään viisi kertaa harrastuksissa viikon aikana?

Summa-frekvenssi-jakaumat

Kun lasketaan frekvenssejä yhteen, saadaan summafrekvenssit. Summafrekvenssit kertovat, kuinka paljon havaintoja on kertynyt tiettyyn muuttujan arvoon mennessä. Tästä syystä summafrekvenssejä kutsutaan **kertymiksi**.

Absoluuttinen **summafrekvenssi** *sf* kuvaa kertymää lukumäärien avulla. Tämä summafrekvenssi saadaan laskemalla (absoluuttisia) frekvenssejä yhteen.

Sisarusten lukumäärä	Frekvenssi *f*	Summafrekvenssi *sf*
0	3	3
1	11	3 + 11 = 14
2	6	14 + 6 = 20
3	2	20 + 2 = 22

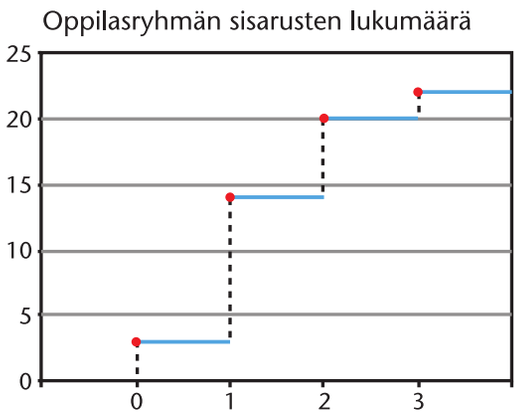
Havaintoaineistossa on **14** henkilöä, joilla on sisaruksia 0–1 kappaletta.

Viimeisen summafrekvenssin pitää olla sama kuin havaintoyksiköiden lukumäärä:
3 + 11 + 6 + 2 = 22

Myös summafrekvenssejä voidaan havainnollistaa kuvaajien avulla. Kuvaajaa piirrettäessä on otettava huomioon, millaista muuttujaa tarkastellaan.

Frekvenssien kuvaaminen s. 168

Kun tarkasteltava muuttuja on diskreetti, kuten esimerkiksi sisarusten lukumäärä, summafrekvenssejä kuvataan **porrasdiagrammilla**. Summafrekvenssit merkitään kuvaajaan pisteillä. Kertymä pysyy samana aina seuraavaan muuttujan arvoon saakka.



Kuvaajassa vaakaviivat yhdistetään katkoviivalla.

Suhteellinen summafrekvenssi *sf* % kuvaa kertymää prosentteina. Suhteellinen summafrekvenssijakauma voidaan muodostaa kahdella eri tavalla:

- lasketaan suhteelliset frekvenssit yhteen
- ilmoitetaan summafrekvenssi prosentteina.

Tarkimmat suhteelliset summafrekvenssit saadaan ilmoittamalla summafrekvenssi prosentteina.

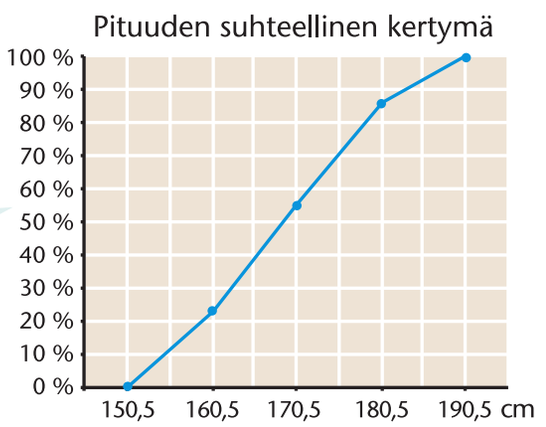
Pituus (cm)	*f*	*f* %	*sf* %	
			Suhteellisten frekvenssien summa	Summafrekvenssi prosentteina
151–160	5	23 %	23 %	$\frac{5}{22} \approx 23\,\%$
161–170	7	32 %	23 % + 32 % = 55 %	$\frac{12}{22} \approx 55\,\%$
171–180	7	32 %	55 % + 32 % = 87 %	$\frac{19}{22} \approx 86\,\%$
181–190	3	14 %	87 % + 14 % = 101 %	100 %

Koska suhteelliset frekvenssit ovat yleensä pyöristettyjä likiarvoja, suhteellisen summafrekvenssin viimeinen arvo voi joskus poiketa jonkin verran 100 %:sta.

Viimeisen suhteellisen summafrekvenssin pitää olla 100 %.

Frekvenssien kuvaaminen s. 168

Kun tarkasteltava muuttuja on jatkuva, kuten esimerkiksi pituus, summafrekvenssejä kuvataan **summakäyrän** avulla. Summakäyrän muodostamiseksi tarvitaan summafrekvenssien lisäksi todelliset luokkarajat. Jakauman kuvaaja alkaa nollatasolta. Summafrekvenssit merkitään kuvaajaan pisteillä luokkien **todellisten ylärajojen** kohdalle ja pisteet yhdistetään murtoviivalla.



Kuvaaja alkaa ensimmäisen luokan todellisesta alarajasta 150,5 cm.

Suhteelliset summafrekvenssipisteet luokkien todellisten ylärajojen kohdalla.

Esimerkki 1

Ravintola Lentävän hanhen ravintoloitsija tutki viikon ajan, kuinka suuria seurueita saapui syömään. Tulokset on koottu viereiseen taulukkoon.

Henkilöiden lukumäärä seurueessa x	Seurueiden lukumäärä f
1	62
2	78
3	63
4	55
5	31

a) Muodosta summafrekvenssijakauma ja suhteellinen summafrekvenssijakauma.
b) Kuinka monta prosenttia seurueista oli korkeintaan kolmen henkilön seurueita?
c) Kuinka monta prosenttia seurueista oli vähintään neljän henkilön seurueita?

Ratkaisu

a) Muodostetaan ensin summafrekvenssijakauma ja tämän avulla suhteellinen summafrekvenssijakauma.

Henkilöiden lukumäärä seurueessa x	Frekvenssi f	Summa-frekvenssi sf	Suhteellinen summafrekvenssi sf %
1	62	62	$\frac{62}{289} = 0{,}214\ldots \approx 21\,\%$
2	78	$62 + 78 = 140$	$\frac{140}{289} = 0{,}484\ldots \approx 48\,\%$
3	63	$140 + 63 = 203$	$0{,}702\ldots \approx 70\,\%$
4	55	$203 + 55 = 258$	$0{,}892\ldots \approx 89\,\%$
5	31	$258 + 31 = 289$	100 %

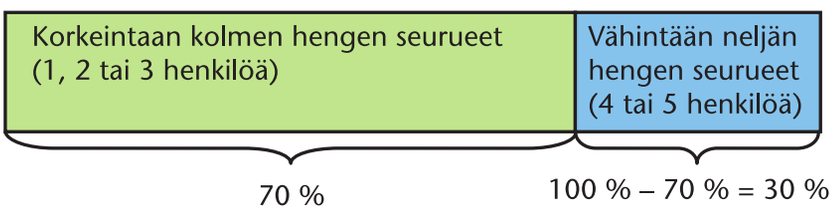
Korkeintaan kolmen hengen seurueita ovat kaikki ne seurueet, joissa henkilöiden lukumäärä on 1, 2 tai 3.

b) Korkeintaan kolmen hengen seurueiden prosenttiosuus nähdään suoraan suhteellisen summafrekvenssin arvosta.

Korkeintaan kolmen henkilön seurueita oli 70 % ravintolaan saapuvista seurueista.

c) Käytetään ratkaisussa apuna b-kohdan tulosta.

Tapa 2
Neljän ja viiden henkilön seurueita oli yhteensä
$55 + 31 = 86$
eli prosentteina
$\frac{86}{289} = 0{,}297\ldots \approx 30\,\%$



Vastaus

b) 70 %
c) 30 %

Mediaani

Tilastolliset tunnusluvut kertovat tutkimusaineiston keskeisiä ominaisuuksia. Tunnusluvuista aiemmin on esitelty moodi, jolla tarkoitetaan aineiston yleisintä arvoa. Moodi voidaan määrittää esimerkiksi frekvenssijakauman perusteella.

Mediaani Md on tunnusluku, joka tarkoittaa suuruusjärjestykseen laitetun aineiston keskimmäistä arvoa. Jos aineisto on pieni, se voidaan helposti laittaa suuruusjärjestykseen ilman jakaumien muodostamista. Tällöin mediaani on se muuttujan arvo, joka on aineistossa keskimmäisenä.

Huom.! Aineisto on ensin laitettava suuruusjärjestykseen.

6, 7, **8**, 9, 9 — Md = 8

Jos keskimmäisiä arvoja on kaksi, mediaani on näiden keskiarvo.

7, 7, **7**, **8**, 8, 9 — $\text{Md} = \frac{7+8}{2} = 7{,}5$

Jos aineistossa on paljon tutkittavia muuttujan arvoja, mediaani kannattaa määrittää suhteellisen summafrekvenssijakauman avulla. Koska mediaani on keskimmäinen arvo, sitä pienempiä arvoja on 50 %. Mediaani on tällöin se muuttujan arvo, jonka suhteellinen summafrekvenssi on ensimmäisenä vähintään 50 %. Luokitellun aineiston mediaanina käytetään luokkakeskusta.

Sisarusten lukumäärä	Suhteellinen summafrekvenssi *sf* %
0	14 %
1	64 %
2	91 %
3	100 %

Suhteellinen summafrekvenssi on ensimmäisenä vähintään 50 %.

Md = 1

Kvartiilit s. 175

Mediaani Md

Mediaani on suuruusjärjestykseen laitetun aineiston keskimmäinen arvo.

ESIMERKKI 2

Valma tutki kirjoittamiensa tekstiviestien merkkien määriä kuukauden ajan. Muodosta tarvittavat frekvenssijakaumat ja määritä aineiston moodi ja mediaani.

Merkkien määrä tekstiviestissä	*f*
0–39	46
40–79	21
80–119	35
120–159	14

RATKAISU

Muodostetaan molemmat summafrekvenssijakaumat.

Merkkien määrä tekstiviestissä	*f*	*sf*	*sf* %
0–39	46	46	$\frac{46}{116} = 0{,}396\ldots \approx 40\ \%$
40–79	21	46 + 21 = 67	$\frac{67}{116} = 0{,}577\ldots \approx 58\ \%$
80–119	35	102	$0{,}879\ldots \approx 88\ \%$
120–159	14	116	100 %

Koska aineiston moodiluokka on se, jonka frekvenssi on suurin, moodiluokka on 0–39 (merkkiä).

Lasketaan moodiluokan luokkakeskus.

$$\text{Mo} = \frac{0+39}{2} = 19{,}5 \text{ (merkkiä)}$$

Taulukon luokkarajat ovat todelliset luokkarajat, koska merkkien määrä on aina kokonaisluku.

Mediaaniluokka on se luokka, jonka suhteellinen summafrekvenssi ylittää ensimmäisen kerran 50 %. Aineiston mediaaniluokka on tällöin 40–79 (merkkiä).

Lasketaan mediaaniluokan luokkakeskus.

$$\text{Md} = \frac{40+79}{2} = 59{,}5 \text{ (merkkiä)}$$

Luokkakeskusten avulla lasketut arvot moodille ja mediaanille voivat olla myös desimaalilukuja.

VASTAUS

Mo = 19,5 merkkiä, Md = 59,5 merkkiä

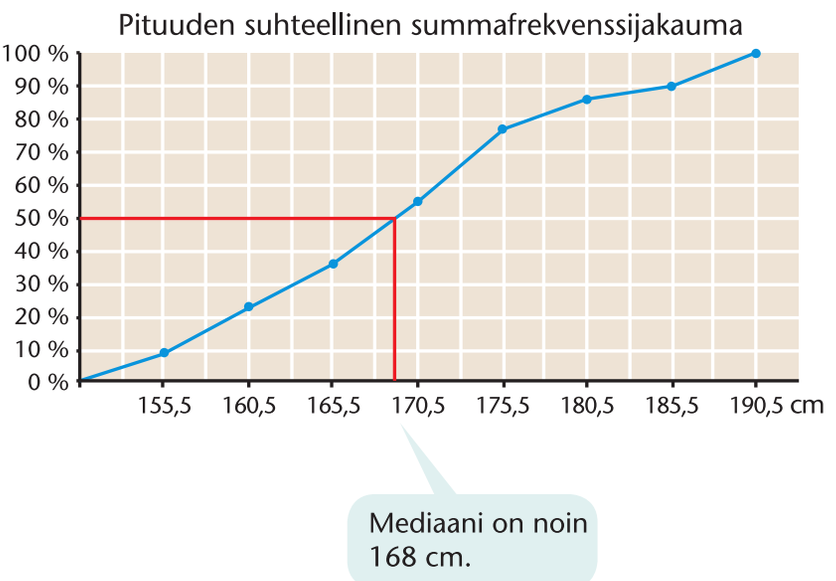
Jatkuvan muuttujan kohdalla aineisto on yleensä luokiteltu. Tällöin suhteellisen summafrekvenssijakauman perusteella voidaan määrittää **mediaaniluokka**.

Luokka	*f*	*f* %	*sf*	*sf* %
151–155	2	9 %	2	9 %
156–160	3	14 %	5	23 %
161–165	3	14 %	8	36 %
166–170	4	18 %	12	55 %
171–175	5	23 %	17	77 %
176–180	2	9 %	19	86 %
181–185	1	5 %	20	91 %
186–190	2	9 %	22	100 %

Mediaaniluokka

Mediaanille voidaan laskea luokitellusta aineistosta myös lukuarvo, mutta sen laskeminen sivuutetaan tällä kurssilla. Luokitellussa aineistossa mediaani on sama kuin luokan luokkakeskus.

Suuntaa antava likiarvo voidaan lukea myös summafrekvenssikäyrältä 50 %:n kohdalta.



ESIMERKKI 3

Oppilaitos järjesti kyselyn opiskelijoilleen, jossa selvitettiin aamuisin koulumatkaan käytettävän ajan keskimääräistä pituutta.

Koulumatkaan käytettävä aika aamuisin (min)	*f*
0–19	74
20–39	154
40–59	129
60–79	82
80–99	12

a) Muodosta suhteellinen summafrekvenssijakauma ja määritä sen avulla mediaani.

b) Määritä aineiston mediaani kuvaajan avulla.

RATKAISU

a) Muodostetaan ensin summafrekvenssi ja tämän avulla suhteellinen summafrekvenssi.

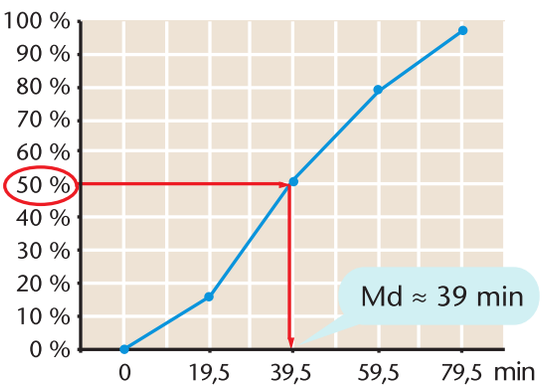
Mediaaniluokka, koska suhteellinen summafrekvenssi ylittää ensimmäisen kerran 50 %.

Aika (min)	*f*	*sf*	*sf* %
0–19	74	74	$\frac{74}{451} \approx 16\,\%$
20–39	154	74 + 154 = 228	$\frac{228}{451} \approx 51\,\%$
40–59	129	357	$\frac{357}{451} \approx 79\,\%$
60–79	82	439	$\frac{439}{451} \approx 97\,\%$
80–99	12	451	$\frac{451}{451} = 100\,\%$

Mediaani on mediaaniluokan luokkakeskus eli

$$\mathrm{Md} = \frac{19{,}5 + 39{,}5}{2} = 29{,}5\,(\mathrm{min})$$

b) Piirretään suhteellisen summafrekvenssin avulla kuvaaja ja luetaan siitä likiarvo mediaanille.



VASTAUS

a) 29,5 min

b) 39 min

Tehtäviä

Summa-frekvenssi-jakaumat

22. Opettaja seurasi viikon ajan, kuinka monta kertaa erään ryhmän oppilaat myöhästyivät aamun ensimmäiseltä tunnilta. Opiskelijat myöhästelivät seuraavasti:

2	0	1	1	3	0	0	2	4	0	1	2	4
1	1	1	0	3	2	2	0	1	1	0	5	2

Laadi aineistosta taulukko, johon lasket kaikki frekvenssit. Vastaa taulukon avulla kysymyksiin.

a) Kuinka monta prosenttia opiskelijoista ei myöhästynyt kertaakaan?
b) Kuinka moni myöhästyi vähintään kahtena aamuna?
c) Kuinka monta prosenttia opiskelijoista myöhästyi enintään kolmena aamuna?

E1

23. Eräässä tutkimuksessa selvitettiin koti talouksissa olevien tietokoneiden lukumääriä.

a) Muodosta aineiston perusteella summafrekvenssijakauma ja suhteellinen summafrekvenssijakauma.
b) Kuinka monta prosenttia kotitalouksista omistaa korkeintaan kolme tietokonetta?
c) Kuinka monta prosenttia kotitalouksista omistaa vähintään neljä tietokonetta?

Tietokoneiden lukumäärä kotitaloudessa	*f*
0	8
1	191
2	462
3	136
4	58
5	6

24. Lukion lyhyen matematiikan äänimäärät erään lukion kevään ylioppilaskirjoituksissa olivat seuraavat:

Äänimäärä	0	2	3	4	5	6	7
Kpl	1	4	7	15	18	12	2

Laske summafrekvenssi ja suhteellinen summafrekvenssi. Vastaa niiden avulla kysymyksiin. Kuinka monta prosenttia opiskelijoista

a) sai enintään neljä ääntä
b) jäi alle neljän äänen
c) sai ainakin kaksi ääntä?

25. Taulukkoon on koottu erään matematiikan ryhmän oppilaiden pituudet (cm).

165	163	168	176	162	178
184	170	167	164	175	168
161	171	163	167	172	161

Täydennä taulukko.

Luokka (cm)	*f*	*f* %	*sf*	*sf* %
160–164				
165–169				
170–174				
175–179				
180–184				

Vastaa frekvenssijakaumien avulla kysymyksiin.

a) Kuinka moni oppilaista on alle 170 cm pitkä?
b) Kuinka monta prosenttia oppilaista on pituudeltaan vähintään 180 cm?
c) Kuinka monta prosenttia oppilaista on pituudeltaan vähintään 165 cm?
d) Kuinka monen oppilaan pituus on alle 175 cm?
e) Kuinka monen oppilaan pituus on vähintään 170 cm, mutta alle 180 cm?
f) Kuinka monta prosenttia oppilaista on pituudeltaan vähintään 165 cm, mutta alle 180 cm?
g) Kuinka moni oppilaista on yli 160 cm?

26. Taulukossa on erään lukion opiskelijoi den preliminäärikokeeseen käyttämän ajan frekvenssijakauma. Muodosta summafrekvenssijakaumat. Kuinka monta prosenttia opiskelijoista

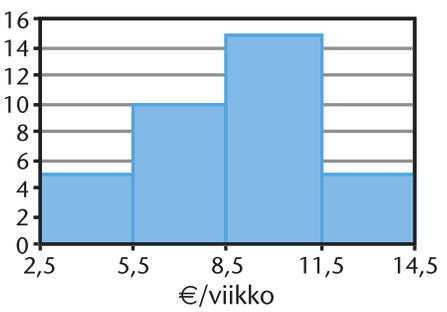
a) viipyi kokeessa alle kaksi tuntia
b) istui kokeessa yli viisi tuntia?

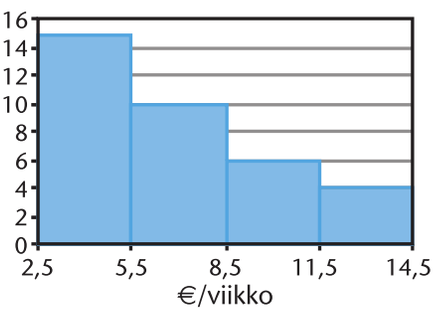
Aika (min)	Opiskelijoita
0–59	2
60–119	3
120–179	7
180–239	28
240–299	25
300–360	11

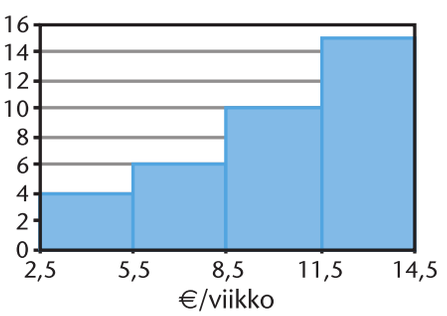
27. Laske taulukosta puuttuvat frekvenssit.

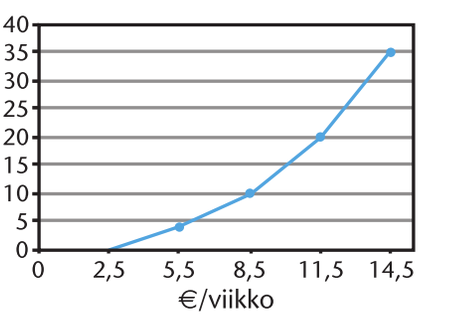
f	*f* %	*sf*	*sf* %
2	10 %		
		6	
			80 %

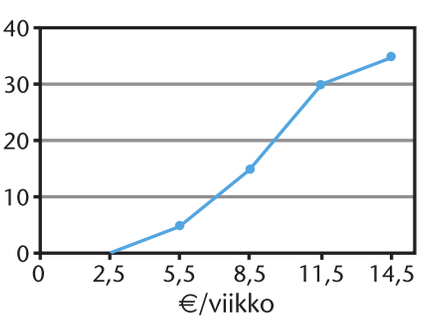
28. Laura teki tutkimusta opiskelijoiden viikkorahoista. Hän keräsi aineistonsa kolmesta eri ryhmästä ja piirsi seuraavat jakaumat. Yhdistä oikea frekvenssijakauma ja summafrekvenssijakauma.

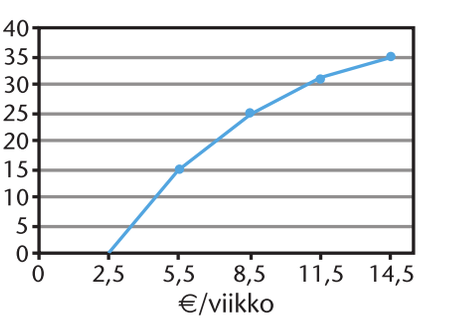












29. Piirrä summafrekvenssijakauman kuvaaja.

a)

Lomamatkojen lukumäärä vuodessa	*sf*
0	8
1	69
2	112
3	142
4	148
5	154
6	165

b)

Pituushypyn tulos (cm)	*sf*
250–299	25
300–349	90
350–399	135
400–449	167
450–499	180
500–549	208
550–599	215

30. Suomalaisten sijoitukset jääkiekon MM-kilpailuissa vuosina 1939–2012 olivat seuraavat:

Sijoitus	1.	2.	3.	4.	5.	6.	7.	9.	13.
Kpl	2	6	3	11	16	8	11	1	1

Muodosta frekvenssi- ja summafrekvenssijakaumat. Havainnollista sopivalla kuvaajalla
a) frekvenssijakaumaa b) summafrekvenssijakaumaa.

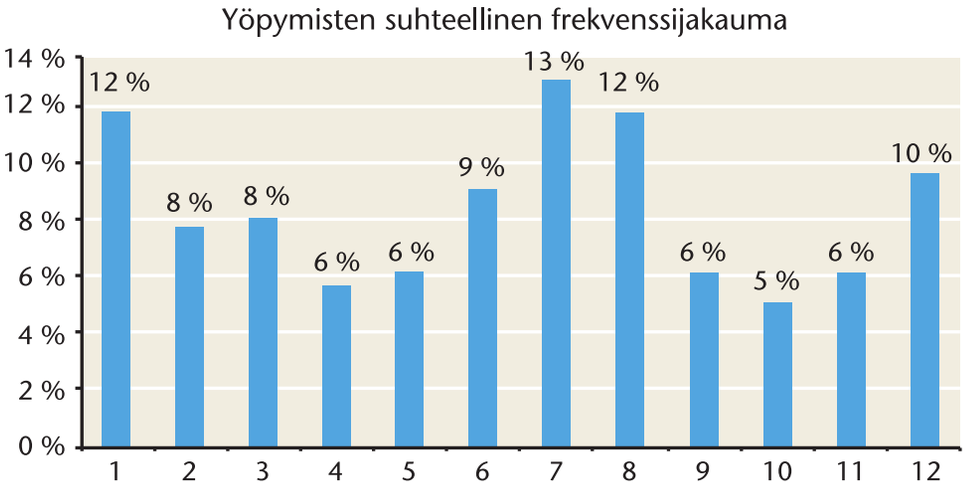
31. Erään suomalaisen teollisuusyrityksen henkilöstön ikärakenne on kuvattuna taulukossa. Laske henkilöstön tiedoista suhteelliset summafrekvenssit. Piirrä aineiston pohjalta
a) histogrammi
b) summakäyrä.

Ikä	Työntekijöiden lukumäärä
21–25	28
26–30	35
31–35	51
36–40	109
41–45	80
46–50	90
51–55	43
56–60	17
61–65	10

Mediaani

32. Minean kurssiarvosanat matematiikan kursseista olivat 9, 8, 8, 7, 9 ja 9. Määritä kurssiarvosanojen moodi ja mediaani.

33. Vuonna 2012 eri Euroopan maista saapuneet vieraat yöpyivät suomalaisissa majoitusliikkeissä yhteensä 4 703 827 kertaa. Nämä majoituskerrat jakaantuivat eri kuukausille kuvan mukaisesti.



Lähde: Stakes

a) Mikä on aineiston moodi? Mitä se tässä aineistossa tarkoittaa?
b) Mikä on aineiston mediaani? Mitä se tässä aineistossa tarkoittaa?
c) Kuinka monta yöpymistä eri Euroopan maista saapuneet vieraat käyttivät joulukuussa?

34. Määritä aineiston moodi ja mediaani.

a)

Hiusten pesujen lukumäärä viikossa	*f*	*sf* %
1	12	13 %
2	32	49 %
3	15	66 %
4	8	74 %
5	10	86 %
6	9	96 %
7	4	100 %

b)

Kuukausipalkka (€)	*f*	*sf*
1500–1999	16	16
2000–2499	42	58
2500–2999	38	96
3000–3499	42	138
3500–3999	31	169
4000–4499	19	188

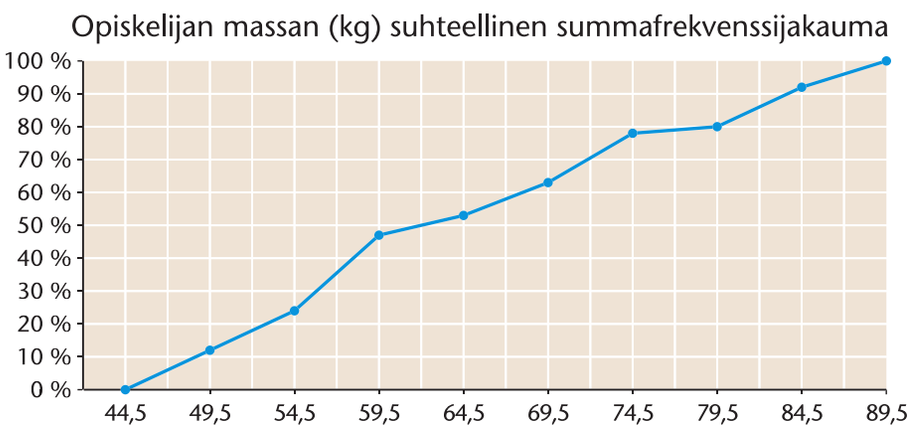
E2 35. Vuonna 2007 Suomessa oli lähes 500 000 kesämökkiä eri puolella maata. Kuntia, joissa kesämökkejä oli alle 1000 kappaletta, oli yhteensä 177. Muodosta tarvittavat frekvenssijakaumat ja määritä aineiston moodi ja mediaani. Mitä moodi ja mediaani tarkoittavat tässä aineistossa?

Kesämökkien lukumäärä kunnassa	*f*
0–99	12
100–199	16
200–299	23
300–399	25
400–499	23
500–599	17
600–699	6
700–799	25
800–899	13
900–999	17

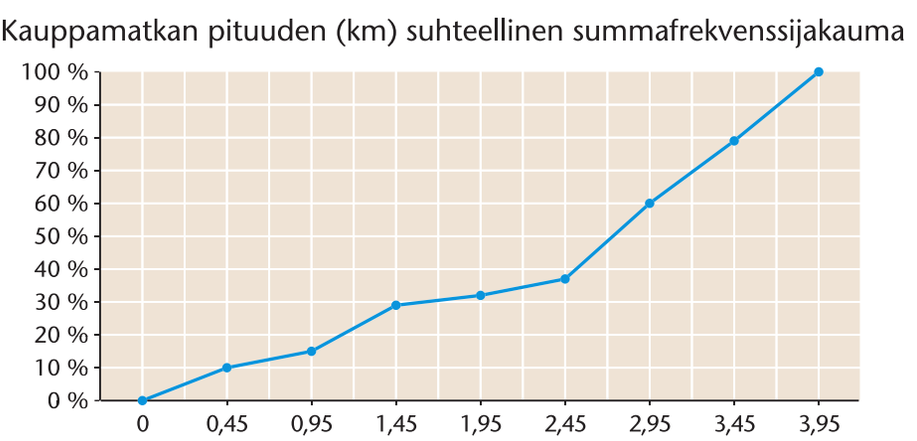
Lähde: Stakes

36. Määritä kuvaajan avulla aineiston mediaani.

a)



b)



37. Lukion päättötodistuksen saaminen edellyttää vähintään 75 kurssin suorittamista. Monet suorittavat enemmän kursseja. Taulukossa on erään lukion abiturienttien kurssimäärien jakauma. Määritä seuraavat tunnusluvut ja selitä, mitä ne tässä aineistossa kuvaavat:
a) mediaaniluokka
b) moodiluokka.

Kurssimäärä	Abiturientteja
75–77	87
78–80	34
81–83	17
84–86	9
87–89	5
90–92	2
yhteensä	154

38. Määritä puhelinlaskun suuruuden mediaani suhteellisen summafrekvenssijakauman kuvaajan avulla.

Puhelinlaskun suuruus (€)	*sf* %
0–9,9	5 %
10,0–19,9	15 %
20,0–29,9	35 %
30,0–39,9	40 %
40,0–49,9	75 %
50,0–59,9	85 %
60,0–69,9	100 %

E3 39. Opiskelijakyselyssä tarkasteltiin oppilaitoksen opiskelijoiden tietokoneen käyttöä. Eräs tutkimuksen muuttujista oli vuorokaudessa tietokoneen käyttöön kuluva aika minuutteina.
a) Muodosta muuttujan suhteellinen summafrekvenssijakauma ja määritä sen avulla mediaani.
b) Määritä muuttujan mediaani kuvaajan avulla.

Tietokoneen käyttöaika vuorokaudessa (min)	*f*
0–29	42
30–59	124
60–89	126
90–119	41
120–149	60
150–179	12
180–209	7

Sekalaisia tehtäviä

40. Golf-amatööri taulukoi viiden vuoden aikana pelatut yhdeksän reiän pelit. Hänen tuloksensa olivat:

Lyöntejä/peli	40	41	42	43	44	45	46
Pelejä	19	2	14	5	9	13	6

Muodosta kaikki frekvenssijakaumat ja vastaa kysymyksiin.

a) Kuinka monta peliä amatööri pelasi alle 45 lyönnin?

b) Kuinka monessa prosentissa peleistä hän selvisi korkeintaan 42 lyönnillä?

c) Kuinka monta prosenttia peleistä hän pelasi yli 43 lyönnin?

41. Piirrä summafrekvenssijakauman kuvaaja.

a)

Kuukaudessa luettujen kirjojen lukumäärä	*f*
0	65
1	30
2	26
3	19
4	39
5	16

b)

100 m juoksuun kuluva aika (s)	*f*
9–11	2
12–14	9
15–17	27
18–20	42
21–23	56

42. Lemmikkieläinkauppa oli lauantaisin auki klo 9–18. Kauppias kirjasi eräänä lauantaina joka tunti ylös yhden tunnin aikana käyneiden asiakkaiden lukumäärät. Asiakasmäärät olivat seuraavat: 2, 11, 6, 14, 15, 8, 7, 17, 4. Määritä asiakkaiden lukumäärän mediaani. Mitä mediaani tässä aineistossa tarkoittaa?

43. Matematiikan tunnilla tutkittiin, kuinka pitkiä yöunia ryhmän opiskelijat keskimäärin nukkuivat.

a) Muodosta muuttujan suhteellinen summafrekvenssijakauma ja määritä sen avulla yöunen pituuden mediaani.

b) Määritä muuttujan mediaani kuvaajan avulla.

Yöunien pituus (min)	*f*
360–389	2
390–419	4
420–449	10
450–479	8
480–509	7
510–539	3
540–569	1

1.3 Keskiarvo

POHDINTA 1

Henrikillä on peruskoulun päästötodistuksessa seuraavat arvosanat lukuaineissa:

Oppiaine	Arvosana
äidinkieli	8
ruotsi	7
englanti	9
matematiikka	9
uskonto	8
fysiikka	9
kemia	8
biologia	7
maantieto	9
historia	7

a) Henrik aikoo hakea lukioon, jonne vaaditaan lukuaineissa keskiarvo 8,2. Pääseekö hän tähän lukioon?

b) Henrik on kiinnostunut luonnontieteistä ja matematiikasta. Siksi hän hakee myös luonnontiedelukioon. Sinne painotetaan päästötodistuksen matematiikan, fysiikan ja kemian arvosanoja seuraavasti:

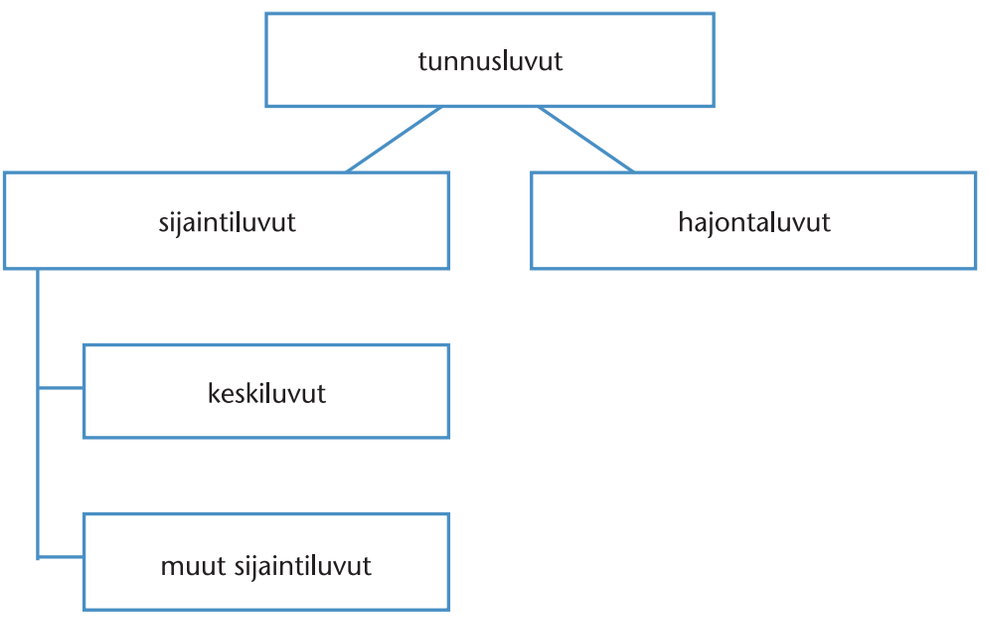
- Matematiikan arvosanaa painotetaan kolminkertaisesti muihin arvosanoihin verrattuna.
- Fysiikan ja kemian arvosanoja painotetaan kaksinkertaisesti muihin arvosanoihin verrattuna.

Luonnontiedelukioon pääsemiseksi on keskiarvo edellä painottaen oltava vähintään 8,8. Pääseekö Henrik luonnontiedelukioon?

Havaintoaineiston kuvaaminen erilaisten frekvenssien ja diagrammien avulla ei yleensä anna riittävän tarkkaa tietoa päätöksenteon tueksi. Tästä syystä havaintoaineistosta lasketaan erilaisia **tunnuslukuja**. Aikaisemmissa luvuissa on jo tutustuttu kahteen tunnuslukuun, moodiin ja mediaaniin.

Tunnusluvut jaetaan kahteen luokkaan: **sijaintilukuihin** ja **hajontalukuihin**. Sijaintiluvut kertovat olennaista tietoa muuttujan arvojen sijainnista: mikä muuttujan arvo on yleisin, mikä keskimmäinen jne. Hajontaluvut, kuten keskihajonta, kertovat, kuinka laajalle alueelle muuttujan arvot ovat hajaantuneet keskiarvosta.

Sijaintiluvut jaetaan **keskilukuihin** ja muihin sijaintilukuihin. Keskiluvut, kuten keskiarvo ja mediaani, kuvaavat muuttujan arvojen keskimääräisiä ominaisuuksia.



Aritmeettinen keskiarvo

Sijaintiluvuista eniten käytetty on **aritmeettinen keskiarvo**. Kun puhutaan keskiarvosta, tarkoitetaan yleensä aritmeettista keskiarvoa. Muita keskiarvoja ovat mm. harmoninen keskiarvo ja geometrinen keskiarvo. Tässä kirjassa käsitellään aritmeettista ja painotettua keskiarvoa.

Aritmeettinen keskiarvo $\bar{x}$ kuvaa keskimääräistä muuttujan arvoa. Se lasketaan jakamalla muuttujan arvojen summa arvojen lukumäärällä.

ESIMERKKI 1

Uuden matkapuhelimiin pelejä valmistavan yrityksen liikevoitot (miljoonia euroja) neljänä ensimmäisenä toimintavuonna on koottu taulukkoon:

Vuosi	Liikevoitto (milj. €)
2010	2,12
2011	2,45
2012	2,80
2013	2,78

a) Laske yrityksen keskimääräinen vuotuinen liikevoitto neljän ensimmäisen vuoden aikana.

b) Vuonna 2014 liikevoiton ennustetaan kasvavan 8,0 % edellisen vuoden voittoon verrattuna. Laske viiden vuoden keskimääräinen liikevoitto. Kuinka monta prosenttia suurempi se olisi verrattuna neljän ensimmäisen vuoden keskiarvoon?

RATKAISU

a) Lasketaan neljän ensimmäisen vuoden liikevoittojen aritmeettinen keskiarvo:

$$\bar{x} = \frac{2{,}12 + 2{,}45 + 2{,}80 + 2{,}78}{4} = 2{,}5375 \approx 2{,}54 \text{ (milj. €)}$$

100 % + 8 % = 108 % = 1,08

b) Vuoden 2014 liikevoitoksi ennustetaan

$$1{,}08 \cdot 2{,}78 = 3{,}0024 \text{ (milj. €)}$$

Viiden vuoden keskiarvo olisi

$$\bar{x} = \frac{2{,}12 + 2{,}45 + 2{,}80 + 2{,}78 + 3{,}0024}{5}$$
$$= 2{,}630... \approx 2{,}63 \text{ (milj. €)}$$

Verrataan viiden vuoden keskiarvoa neljän vuoden keskiarvoon:

$$\frac{2{,}630...}{2{,}5375} = 1{,}03664... = 103{,}664... \%$$

Voitto on siis 103,664... % − 100 % = 3,664... % ≈ 3,66 % suurempi.

VASTAUS

a) 2,54 milj. € b) 2,63 milj. €, kasvua 3,66 %

Muuttujan arvoja kuvataan erilaisissa laskukaavoissa merkinnällä x_i. Alaindeksin i tarkoitus on erottaa luvut toisistaan. Jos esimerkiksi kolmen oppilaan kurssiarvosanat ovat 6, 7 ja 9, niin voidaan merkitä

$$x_1 = 6 \qquad x_2 = 7 \qquad x_3 = 9$$

Tilastotieteessä lasketaan usein erilaisten lukujen summia. Koska yhteenlaskettavia on yleensä paljon, kaikkia laskutoimituksia on hankala kirjoittaa näkyviin. Tästä syystä käyttöön on otettu lyhyempiä merkintätapoja. Lukujen summaa merkitään lyhyesti summamerkillä $\sum$ (sigma).

Arvosanojen $x_1 = 6$, $x_2 = 7$ ja $x_3 = 9$ summa voidaan merkitä summamerkin avulla seuraavasti:

$$\sum_{i=1}^{3} x_i = x_1 + x_2 + x_3 = 6 + 7 + 9$$

Indeksi i saa arvot 1, 2 ja 3. Pienin arvo merkitään summamerkin alapuolelle ja suurin arvo summamerkin yläpuolelle.

Summamerkkiä käyttäen näiden kolmen luvun keskiarvo merkitään

$$\bar{x} = \frac{\sum_{i=1}^{3} x_i}{3} = \frac{6+7+9}{3} \approx 7{,}3$$

Yleisesti summaa, jonka yhteenlaskettavat ovat x_1, x_2, x_3,…, x_n merkitään

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + \ldots + x_n$$

n on yhteenlaskettavien lukumäärä

Keskiarvon laskukaava yleisessä muodossa on

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Yleensä havaintoaineistossa sama muuttujan arvo esiintyy useita kertoja. Jos esimerkiksi arvosana $x_1 = 6$ esiintyy kolme kertaa, arvosana $x_2 = 7$ kaksi kertaa ja arvosana $x_3 = 9$ neljä kertaa, arvosanojen esiintymiskertoja merkitään frekvenssien f_i avulla seuraavasti:

$$f_1 = 3 \qquad f_2 = 2 \qquad f_3 = 4$$

x_i	f_i
6	3
7	2
9	4

Arvosanojen summa on tällöin $\sum_{i=1}^{3} f_i x_i = 3 \cdot 6 + 2 \cdot 7 + 4 \cdot 9$.

Summamerkkiä käyttäen keskiarvo merkitään seuraavasti:

$$\bar{x} = \frac{\sum_{i=1}^{3} f_i x_i}{9} = \frac{3 \cdot 6 + 2 \cdot 7 + 4 \cdot 9}{9} \approx 7{,}6$$

Osa tunnusluvuista saadaan laskettua myös laskimen tilastotoimintojen avulla. Monissa laskimissa on mahdollista syöttää aineistoja laskimen muistiin. Aineiston syöttämisen jälkeen esimerkiksi aritmeettinen keskiarvo saadaan laskimesta valmiiksi ohjelmoidun laskukaavan avulla. Tutustu oman laskimesi tilastotoimintoihin laskimen ohjekirjan avulla.

Aritmeettinen keskiarvo $\bar{x}$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{n}$$

- n on havaintoyksiköiden lukumäärä.
- k on muuttujan arvojen tai luokkien lukumäärä.
- x_i on muuttujan arvo tai luokan luokkakeskus.
- f_i on muuttuja arvon tai luokan frekvenssi.

ESIMERKKI 2

Oppilaskunta suunnittelee virvoitusjuoma-automaatin hankkimista koululle. Oppilaskunta haastattelee 22 opiskelijaa selvittääkseen, kuinka monta tölkkiä viikossa juoman kulutus olisi. Automaatin hankkiminen kannattaa, jos kulutus opiskelijaa kohti on viikossa vähintään kaksi tölkkiä. Haastattelun tulokset (tölkkiä/oppilas) on koottu seuraavaan taulukkoon.

0	3	2	2	4	1	1	1	5	2	1
2	5	0	1	2	1	4	2	1	4	3

Saako koulu virvoitusjuoma-automaatin?

RATKAISU

Muuttuja ”viikottaisten juoma-tölkkien lukumäärä” saa vain arvoja 0–5. Muodostetaan tölkkien lukumäärän frekvenssijakauma, jotta saadaan selville, montako kutakin muuttujan arvoa on.

Tölkkien lukumäärä (x_i)	Frekvenssi f_i
0	2
1	7
2	6
3	2
4	3
5	2

Muuttujan arvojen summa on

$$2 \cdot 0 + 7 \cdot 1 + 6 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + 2 \cdot 5 = 47$$

frekvenssi f_i — muuttujan arvo x_i

Yhteenlaskettavat ovat siis muotoa $f_i \cdot x_i$.

Kyselyyn osallistui 22 opiskelijaa eli havaintoyksiköiden lukumäärä n on 22.

Koska muuttujan eri arvoja on kuusi, indeksi i saa arvot 1–6.

$$\bar{x} = \frac{\sum_{i=1}^{6} f_i x_i}{22} = \frac{2 \cdot 0 + 7 \cdot 1 + 6 \cdot 2 + 2 \cdot 3 + 3 \cdot 4 + 2 \cdot 5}{22} = \frac{47}{22} = 2{,}136\ldots \approx 2{,}1$$

Tutkimuksen perusteella opiskelija juo keskimäärin 2,1 tölkkiä viikossa eli automaatti voidaan hankkia koululle.

VASTAUS

Saa, koska juomien kulutus olisi 2,1 tölkkiä/opiskelija.

ESIMERKKI 3

Työttömyysaste nousi ennätyskorkealle Euroopan Unionin alueella vuoden 2012 aikana. Alla olevaan taulukkoon on koottu eri EU-maiden työttömyysasteita vuodelta 2012. (Lähde: Eurostat)

Työttömyysaste (%)	f
5–9	12
10–14	8
15–19	4
20–24	0
25–29	2

Määritä aineiston avulla keskimääräinen työttömyysaste Euroopan Unionissa.

RATKAISU

Kun lasketaan tunnuslukuja luokitellulle aineistolle, muuttujan arvoina x_i käytetään luokkien luokkakeskuksia. Määritetään luokkakeskukset todellisten rajojen avulla.

Työttömyysaste (%)	Tod. alara	Tod. yläraja	Luokka-keskus	f
5–9	4,5	9,5	7	12
10–14	9,5	14,5	12	8
15–19	14,5	19,5	17	4
20–24	19,5	24,5	22	0
25–29	24,5	29,5	27	2

$\frac{4{,}5+9{,}5}{2}=7$

Aineiston havaintoyksiköiden määrä n on 26. Koska luokkien lukumäärä on viisi, i saa arvot välillä 1–5. Aineiston keskiarvo on siis

$$\bar{x}=\frac{\sum_{i=1}^{k} f_i x_i}{n}=\frac{\sum_{i=1}^{5} f_i x_i}{26}=\frac{12\cdot 7+8\cdot 12+4\cdot 17+0\cdot 22+2\cdot 27}{26}$$

$$=11{,}615\ldots\approx 11{,}6\,(\%)$$

VASTAUS

Euroopan Unionin työttömyysaste oli 11,6 %.

ESIMERKKI 4

Urheiluseurojen välisessä mäkihyppykilpailussa annettiin pisteitä sekä pituudesta että tyylistä. Tyylipisteitä antoivat viisi tuomaria asteikolla 0–20 puolen pisteen tarkkuudella. Hyppääjän lopputulos tyylissä saatiin laskemalla viiden tuomarin antamien pisteiden keskiarvo yhden desimaalin tarkkuudella. Heikki sai kolmelta ensimmäiseltä tuomarilta pisteet 18, 16 ja 17,5.

a) Neljäs tuomari antoi Heikille vain 14 pistettä, mutta keskiarvoksi tuli kuitenkin 16,8. Kuinka paljon viides tuomari antoi?

b) Kaksi viimeistä tuomaria antoivat täsmälleen saman pistemäärän, jolloin keskiarvoksi tuli 16,5. Määritä kahden viimeisen tuomarin antamat pisteet.

RATKAISU

a) Merkitään viidennen tuomarin antamaa pistemäärää kirjaimella x. Tuomarien antamien pisteiden summa on

$$18 + 16 + 17{,}5 + 14 + x = 65{,}5 + x$$

Koska pisteiden keskiarvo on 16,8, niin

$$\begin{aligned} \frac{65{,}5 + x}{5} &= 16{,}8 \qquad \big| \cdot 5 \\ 65{,}5 + x &= 84 \\ x &= 18{,}5 \end{aligned}$$

Viides tuomari antoi 18,5 pistettä.

b) Merkitään yhden tuomarin antamia pisteitä kirjaimella x, jolloin pisteiden summa on

$$18 + 16 + 17{,}5 + x + x = 51{,}5 + 2x$$

Koska keskiarvo on 16,5, niin

$$\begin{aligned} \frac{51{,}5 + 2x}{5} &= 16{,}5 \qquad \big| \cdot 5 \\ 51{,}5 + 2x &= 82{,}5 \\ 2x &= 31 \qquad \big| : 2 \\ x &= 15{,}5 \end{aligned}$$

Molemmat tuomarit antoivat 15,5 pistettä.

VASTAUS

a) 18,5 pistettä b) 15,5 pistettä

Painotettu keskiarvo

On tilanteita, joissa kaikki havaintoaineiston luvut eivät ole yhtä merkityksellisiä. Tällöin ei ole järkevää laskea lukujen aritmeettista keskiarvoa, vaan **painotettu keskiarvo**. Siinä jokaisella luvulla on oma **painokertoimensa**. Painokerroin on positiivinen luku, joka on sitä suurempi, mitä merkityksellisempi luku on.

Painotettu keskiarvo lasketaan seuraavasti:

1. Kerrotaan jokainen muuttujan arvo x_i omalla painokertoimellaan.
2. Lasketaan saadut tulot yhteen.
3. Jaetaan summa painokertoimien summalla.

Esimerkki 5

Yliopisto-opiskelija on suorittanut matematiikan perusopinnot, joiden laajuus on 25 opintopistettä (op). Hän on saanut kursseista seuraavat arvosanat:

Kurssi	Laajuus (op)	Arvosana
johdantokurssi	4	5
symbolinen laskenta	2	4
differentiaalilaskenta	8	3
lineaarialgebra	5	3
todennäköisyyslaskenta	6	2

Perusopintojen keskiarvoa laskettaessa jokaista saatua arvosanaa painotetaan sen mukaan, mikä on kurssin laajuus opintopisteinä. Laske perusopintojen keskiarvo yhden desimaalin tarkkuudella.

Ratkaisu

Painokertoimena on kurssin laajuus. Kun kerrotaan kurssiarvosanat kurssin laajuudella, summaksi saadaan

$$4 \cdot 5 + 2 \cdot 4 + 8 \cdot 3 + 5 \cdot 3 + 6 \cdot 2 = 79$$

Painokerrointen summa on $4 + 2 + 8 + 5 + 6 = 25$.

Painotettu keskiarvo on siis $\frac{79}{25} = 3{,}16 \approx 3{,}2$.

Vastaus

3,2

Tehtäviä

Aritmeettinen keskiarvo

44. Taulukkoon on koottu saksalaisten ja ruotsalaisten turistien majoittumisia Suomessa. Laske tilaston avulla, minkä verran Suomessa keskimäärin yöpyy vuodessa
a) ruotsalaisia turisteja b) saksalaisia turisteja.

Asuinmaa	1998	1999	2000	2001	2002
Ruotsi	570 640	581 477	643 000	628 473	598 851
Saksa	458 235	483 507	477 924	489 066	513 703

Lähde: Tilastokeskus

E1 45. Työttömyysasteet Suomessa vuosina 2007–2011 olivat:

Vuosi	2007	2008	2009	2010	2011
Työttömyysaste (%)	8,9	8,1	9,8	11,6	11,3

a) Laske tämän viisivuotiskauden keskimääräinen työttömyysaste.
b) Vuonna 2011 eräät taloustieteilijät ennustivat Suomen työttömyysasteen laskevan vuonna 2012 peräti 10 %. Jos tämä ennuste olisi pitänyt paikkansa, mikä olisi ollut kauden 2007–2012 keskimääräinen työttömyysaste?

E2 46. Tikkakilpailussa jokainen pelaaja heittää 18 tikkaa. Se pelaaja, jolla heittojen keskiarvo on suurin, voittaa kilpailun. Anssin heittotulokset on koottu taulukkoon.

8	6	6	8	9	9	5	10	3
9	6	8	7	4	5	9	10	9

Kilpailua johtaa Aku, jonka heittojen keskiarvo on 8,1. Voittaako Anssi Akun?

47. Opettaja kokosi tilaston matematiikan kurssin poissaoloista. Oppilaskohtaiset poissaolotunnit on koottu taulukkoon.

Poissaolotunteja	0	1	2	3	4	5	6	7	8	9
Opiskelijoita	8	5	6	2	5	2	2	0	1	1

a) Määritä poissaolojen mediaani ja moodi.
b) Kuinka monta tuntia opiskelija keskimäärin oli poissa kurssilta?

E3 48. Taulukkoon on koottu Afrikan valtioiden vuotuisia väestönkasvuprosentteja. Määritä annettujen tietojen avulla keskimääräinen vuotuinen kasvuprosentti.

Kasvuprosentti	Valtioiden lukumäärä
0,1–1,0	2
1,1–2,0	5
2,1–3,0	25

Kasvuprosentti	Valtioiden lukumäärä
3,1–4,0	18
4,1–5,0	1
5,1–6,0	1

49. Taulukossa on Suomen väestön ikärakenne vuonna 2000. Laske aineiston pohjalta väestön keski-ikä, moodiluokka ja mediaaniluokka. Stakesin tutkimusten mukaan keski-ikä vuonna 2011 oli 41,7 vuotta. Mieti syitä, joiden vuoksi tuloksesi eroaa tästä arvosta.

Ikä	Henkilöitä
0–4	291 275
5–14	645 058
15–64	3 467 584
65–74	436 789
75–84	262 014
85–104	78 395
yhteensä	5 181 115

Lähde: Stakes

50. Eräissä OECD-maissa verojen osuus bruttokansantuotteesta (%) vuonna 2001 oli seuraava:

46,3 53,2 49,0 36,4 39,9 45,3 45,7 42,4 45,4
37,4 29,2 41,8 35,2 40,8 44,9 34,8 34,5 35,2

Lähde: Verohallinto

Luokittele aineisto ja tarkastele OECD-maiden veroastetta eri sijaintilukujen avulla.

E4 51. Kukkarossa on kahden euron, yhden euron ja 50 sentin kolikot. Kolikoiden massat ovat:

8,50 g

7,50 g

7,80 g

a) Kun kukkaroon lisätään yksi 10 sentin kolikko, kolikoiden massojen keskiarvoksi saadaan 6,975 g. Määritä 10 sentin kolikon massa.
b) Kukkaroon lisätään kolmen alussa olevan rahan joukkoon kaksi samanarvoista kolikkoa, jolloin kukkaron rahojen keskimääräinen massa on 7,76 g. Kuinka paljon rahaa kukkarossa on lisäyksen jälkeen?

52. Matematiikan päättötodistuksen arvosana määräytyy pääsääntöisesti aritmeettisesta keskiarvosta. Emman viiden ensimmäisen kurssin arvosanat ovat 8, 6, 7, 8 ja 6.
a) Mikä arvosana hänen on saatava viimeisestä kurssista, jotta loppuarvosana pyöristettäisiin 8:aan?
b) Mikä viimeisen arvosanan pitää vähintään olla, jotta loppuarvosanaksi tulee 7?

53. Erään luvun ja sen neliön keskiarvo on 0,72. Määritä kyseiset luvut.

54. Vuosina 1985–2000 pankkien keskimääräinen liikevoitto oli –29 miljoonaa euroa. Vuosina 1986 ja 1987 liikevoitot olivat yhtä suuret. Laske näiden vuosien liikevoitto.

Vuosi	**1985**	**1986**	**1987**	**1988**	**1989**	**1990**	**1991**	**1992**
milj. (€)	22			74	35	54	–177	–578
Vuosi	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**
milj. (€)	–308	–246	–79	60	136	112	152	219

Lähde: Tilastokeskus

E5

Painotettu keskiarvo

55. Englannin kurssin loppuarvosanaan otettiin huomioon kaksi ainetta, kuuntelukokeen tulos, tekstikokeen tulos ja rakennekokeen tulos. Eri osa-alueiden arvosanoja painotettiin sen mukaan, mikä oli tehtävän suhteellinen osuus prosentteina. Elsa sai kurssin osasuorituksista seuraavat arvosanat:

Osasuoritus	%	Arvosana
aine 1	20	8,5
aine 2	20	8
kuuntelukoe	15	9
tekstikoe	15	7,5
rakennekoe	30	9,5

Laske Elsan tulosten painotettu keskiarvo.

56. Henkilöauton polttoaineen keskikulutusta tutkittiin erilaisissa olosuhteissa tietty aika. Kulutus määritettiin painotettuna keskiarvona siten, että ajoon käytetty aika oli kunkin kulutusmäärän painokertoimena. Määritä tietojen avulla tämän auton keskikulutus.

Olosuhteet	Aika (h)	Kulutus (dm^3/100 km)
tavallinen maantieajo	4	5,80
moottoritieajo	5	6,90
kaupunkiajo	3	8,10
ajo talviolosuhteissa	2	7,20

57. Luonnossa esiintyy kuutta seleenillä isotooppia. Taulukkoon on merkitty kunkin isotoopin suhteellinen runsaus.

Isotooppi	Suhteellinen atomimassa	Runsaus (%)
Se-74	73,922475	0,9
Se-76	75,919212	9,0
Se-77	76,919912	7,6
Se-78	77,917304	23,5
Se-80	79,916520	49,6
Se-82	81,916698	9,4

Seleenin suhteellinen atomimassa lasketaan eri isotooppien atomimassojen keskiarvona ottaen huomioon niiden prosentuaaliset osuudet. Laske seleenin suhteellinen atomimassa.

Sekalaisia tehtäviä

58. Eräällä vesikasvilajilla kukan terälehtien määrä vaihtelee neljän ja kahdeksan lehden välillä. Suolta kerätyistä näytteistä laadittiin taulukko.

Terälehtiä	4	5	6	7	8
f	26	32	33	11	6

a) Määritä terälehtien määrän moodi ja mediaani.
b) Kuinka monta terälehteä kukassa keskimäärin oli?

59. Eräässä matematiikankokeessa arvosanojen jakauma oli seuraava:

4	5	6	7	8	9	10
1,3 %	9,8 %	15,8 %	20,3 %	23,3 %	23,4 %	6,1 %

Laske arvosanojen keskiarvo. [S2001, 5]

60. Jääkiekkojoukkue JYP:n pelaajien pistepörssin pisteet kaudella 2002–2003 olivat seuraavat:

46	34	33	32	32
26	20	19	15	14
13	11	8	7	7
5	5	5	4	2
1	0	0	0	0
0	0	0	0	0

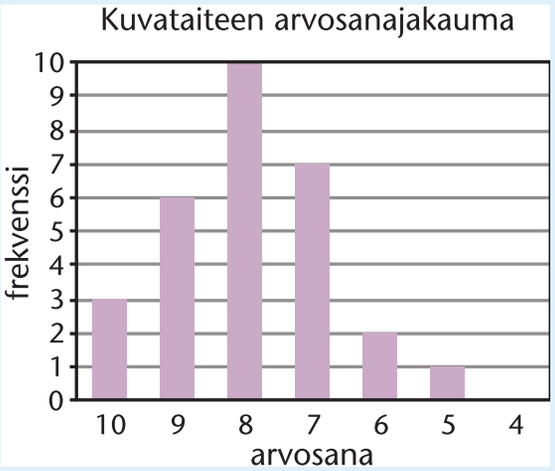
Luokittele pistemäärät ja analysoi pelaajien saamia pistemääriä eri sijaintilukujen avulla.

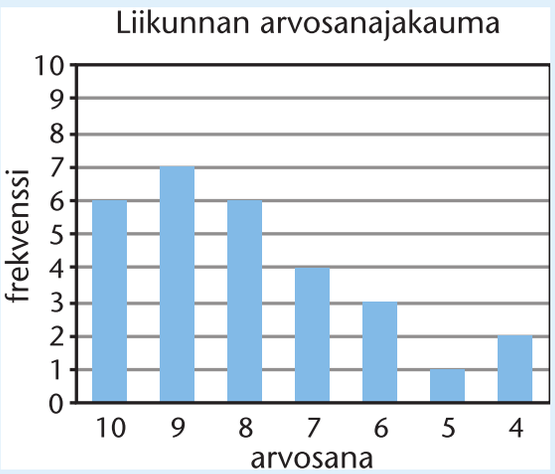
61. Viiden jalkapalloilijan yhteispaino on 425 kg, ja kymmenen balettitanssijan keskimääräinen paino 52 kg. Mikä on kaikkien viidentoista keskimääräinen paino? [K1998, 2]

62. Tutustu jonkin urheilujoukkueen kotisivuihin. Laadi sopivasta aiheesta (kauden aikana tehdyt maalit, kerätyt pisteet, heitetyt korit tms.) tilasto ja analysoi sitä sijaintilukujen avulla.

1.4 Keskihajonta

POHDINTA 1 Erään ryhmän viimeisimmät arvosanat kuvataiteessa ja liikunnassa olivat seuraavat:



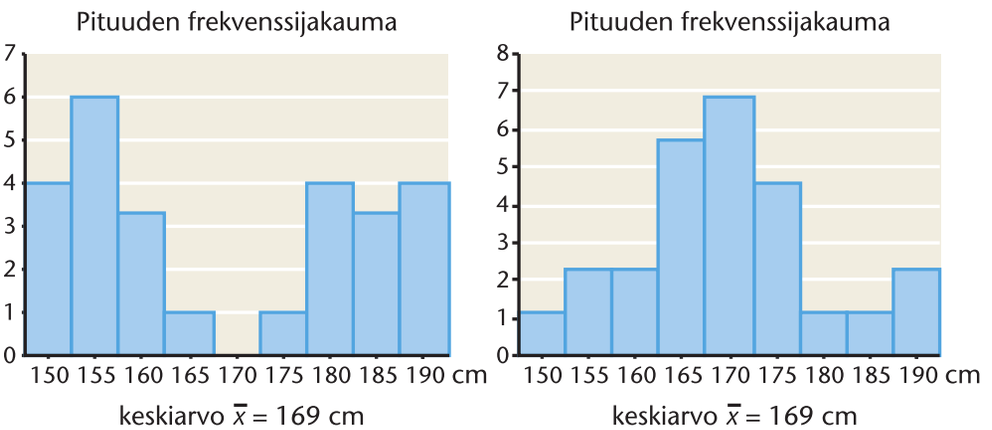


a) Voiko kuvaajien perusteella sanoa, kummassa aineessa tulokset olivat paremmat?

b) Arvioi kuvaajien avulla, kummassa aineessa arvosanojen hajonta oli suurempaa. Perustele vastauksesi.

c) Ville osallistui molemmille kursseille ja sai kummastakin arvosanaksi 7. Kummassa aineessa Ville menestyi suhteessa paremmin? Perustele vastauksesi.

Pelkät keskiluvut eivät riitä kuvaamaan jakaumia. Esimerkiksi kahdella jakaumalla voi olla sama keskiarvo, vaikka ne näyttävät aivan erilaisilta.



Hajontaluvut ovat tunnuslukuja, joilla kuvataan muuttujan arvojen jakaantumista keskiarvon ympärille. Ne kertovat siis, kuinka lähellä keskiarvoa muuttujan arvot ovat.

Erään oppilaan pakollisten kurssien arvosanat olivat 7, 9, 8, 5, 6 ja 7. Arvosanojen keskiarvo on

$$\bar{x} = \frac{7+9+8+5+6+7}{6} = 7$$

Tarkastellaan, kuinka paljon kunkin kurssin arvosana poikkeaa keskiarvosta.

Arvosana x_i	Poikkeama keskiarvosta $x_i - \bar{x}$
7	7 – 7 = 0
9	9 – 7 = 2
8	8 – 7 = 1
5	5 – 7 = –2
6	6 – 7 = –1
7	7 – 7 = 0

Taulukkoon on laskettu muuttujan poikkeamat keskiarvosta $(x_i - \overline{x})$. Kun poikkeamat lasketaan yhteen, huomataan summan olevan nolla.

$$0 + 2 + 1 + (-2) + (-1) + 0 = 0$$

Koska muuttujat kuitenkin poikkeavat keskiarvostaan, poikkeamien summa ei kuvaa muuttujien arvojen hajaantumista.

Poikkeamien itseisarvojen $|x_i - \overline{x}|$ summa on nollasta poikkeava, jos kaikki arvot x_i eivät ole samoja. Poikkeamien itseisarvojen keskiarvoa kutsutaan **keskipoikkeamaksi**. Keskipoikkeama ei kuitenkaan ole niin käyttökelpoinen kuin poikkeamien neliöihin perustuva tunnusluku. Korotetaan siis kaikki poikkeamat toiseen potenssiin.

Arvosana x_i	Poikkeama keskiarvosta $x_i - \overline{x}$	$(x_i - \overline{x})^2$
7	$7 - 7 = 0$	$0^2 = 0$
9	$9 - 7 = 2$	$2^2 = 4$
8	$8 - 7 = 1$	$1^2 = 1$
5	$5 - 7 = -2$	$(-2)^2 = 4$
6	$6 - 7 = -1$	$(-1)^2 = 1$
7	$7 - 7 = 0$	$0^2 = 0$

Lasketaan poikkeamien neliöiden keskiarvo.

$$\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} = \frac{0+4+1+4+1+0}{6} = 1{,}666...$$

Edellä laskettua poikkeamien neliöiden keskiarvoa kutsutaan **varianssiksi** s^2. Varianssin yksikkö ei ole sama kuin muuttujan yksikkö. Jos esimerkiksi alkuperäisen muuttujan yksikkö on senttimetri (cm), varianssin yksikkö on alkuperäisen yksikön neliö eli cm^2. Tästä syystä hajonnan tulkitseminen varianssin avulla on hankalaa.

Kun varianssista otetaan neliöjuuri, saadaan **keskihajonta *s***, jonka yksikkö on sama kuin alkuperäisen muuttujan yksikkö.

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}$$

Arvosanajakauman keskihajonta on $s = \sqrt{1{,}666...} = 1{,}290... \approx 1{,}3$.

Keskihajontaa laskettaessa summa jaetaan yleensä lausekkeella $n - 1$. Tälle löytyy perustelu tilastotieteen teoriasta, johon tällä kurssilla ei ole mahdollisuutta syventyä. Jos jakajana on luku $n - 1$, keskihajontaa kutsutaan **otoskeskihajonnaksi**. Jos tarkastellaan koko perusjoukkoa, keskihajonnan kaavassa jakajana on luku n.

Useimmissa laskimissa on erikseen toiminnot kumpaakin vaihtoehtoa varten. Kun havaintoaineiston koko n on riittävän suuri, ei ole käytännössä merkitystä, jaetaanko luvulla $n - 1$ vai luvulla n. Tässä kirjassa käytetään jakajana lukua $n - 1$.

Keskihajonta *s*

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \qquad s = \sqrt{\frac{\sum_{i=1}^{k} f_i(x_i - \bar{x})^2}{n-1}}$$

- $\bar{x}$ on keskiarvo.
- k on muuttujan arvojen tai luokkien lukumäärä.
- x_i on muuttujan arvo tai luokan luokkakeskus.
- f_i on muuttujan arvon tai luokan frekvenssi.
- n on havaintoyksiköiden lukumäärä.

ESIMERKKI 1 Venla mittasi ulkolämpötilan joka ilta kello 19.00 yhden viikon ajan. Laske ulkolämpötilojen keskihajonta.

Päivä	1	2	3	4	5	6	7
Lämpötila (°C)	9	5	7	10	12	9	6

RATKAISU Lämpötilojen keskiarvo on

$$\bar{x} = \frac{9+5+7+10+12+9+6}{7} = 8{,}285\ldots\,(°\text{C})$$

Keskihajonnan voi laskea myös laskimen tilastotoiminnolla.

Keskihajonnan laskemista varten muodostetaan taulukko, johon merkitään muuttujan arvojen poikkeamat keskiarvosta ja poikkeamien neliöt.

Päivä	Lämpötila (°C)	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	9	0,714…	0,510…
2	5	–3,285…	10,795…
3	7	–1,285…	1,653…
4	10	1,714…	2,938…
5	12	3,714…	13,795…
6	9	0,714…	0,510…
7	6	–2,285…	5,224…
yhteensä			35,428…

Laske mahdollisimman tarkoilla arvoilla. Näkyviin voi kirjoittaa vain muutaman desimaalin.

Keskihajonta on

$$s = \sqrt{\frac{\sum_{i=1}^{7}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{35{,}428}{7-1}} = 2{,}429\ldots \approx 2{,}4\,(°\text{C})$$

VASTAUS Ulkolämpötilojen keskihajonta on 2,4 °C.

ESIMERKKI 2

Atte ja Leevi harrastavat pituushyppyä. Kilpailukauden aikana hypätyt tulokset on koottu alla oleviin taulukoihin.

Atte	
Pituus	***f***
600–609	8
610–619	14
620–629	19
630–639	7
640–649	2

Leevi	
Pituus	***f***
600–609	16
610–619	5
620–629	8
630–639	15
640–649	6

a) Kumpi hyppää keskimäärin pidempiä hyppyjä?
b) Kumpi on tasaisempi hyppääjä?

RATKAISU

a) Lasketaan aineistojen keskiarvot. Niiden laskemiseksi on määritettävä luokkien luokkakeskukset:

Pituus	Luokkakeskus
600–609	604,5
610–619	614,5
620–629	624,5
630–639	634,5
640–649	644,5

Luokkakeskus on todellisten luokkarajojen keskiarvo:
$\frac{599{,}5+609{,}5}{2}=604{,}5$

Aten hyppyjen keskiarvo on

$$\bar{x}=\frac{\sum_{i=1}^{k} f_i x_i}{n}=\frac{8\cdot 604{,}5+14\cdot 614{,}5+19\cdot 624{,}5+7\cdot 634{,}5+2\cdot 644{,}5}{50}$$
$$=620{,}7$$

Leevin hyppyjen keskiarvo on

$$\bar{x}=\frac{\sum_{i=1}^{k} f_i x_i}{n}=\frac{16\cdot 604{,}5+5\cdot 614{,}5+8\cdot 624{,}5+15\cdot 634{,}5+6\cdot 644{,}5}{50}$$
$$=622{,}5$$

Koska Leevin hyppyjen keskiarvo on suurempi, niin Leevin hypyt ovat keskimäärin pidempiä.

b) Syötetään aineistot laskimeen luokkakeskuksia käyttäen. Keskihajonnat saadaan tilastotoimintojen avulla:

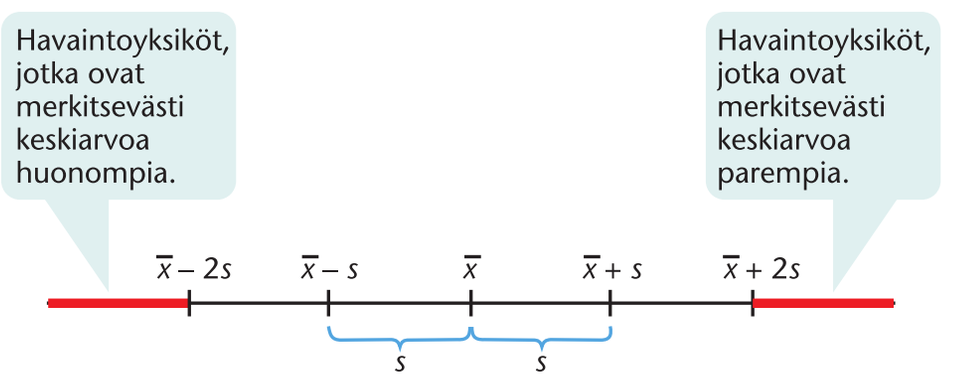
Keskihajonnan voi laskea myös taulukoimalla kuten esimerkissä 1.

Atte: $s \approx 10{,}5$ Leevi: $s \approx 14{,}7$

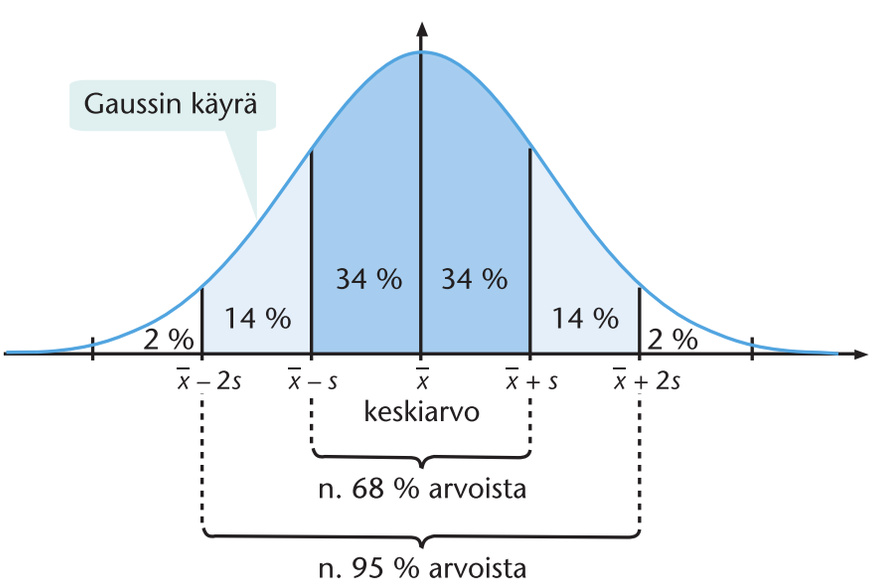
Aten tulosten keskihajonta on pienempi, joten Atte on tasaisempi hyppääjä.

VASTAUS a) Leevi b) Atte

Keskihajonta on hajontaluvuista yleisin ja tärkein. Jos keskihajonta s on pieni, havaintoaineisto on keskittynyt keskiarvon ympärille. Jos muuttujan arvo on vähintään kahden keskihajonnan päässä keskiarvosta, se poikkeaa keskiarvosta merkitsevästi.



Parhaiten keskihajonta on tulkittavissa, kun aineisto on riittävän suuri ja muuttuja noudattaa edes likimain normaalijakaumaa eli Gaussin jakaumaa. Silloin noin 68 % arvoista sijaitsee korkeintaan yhden keskihajonnan päässä keskiarvosta ja noin 95 % kahden keskihajonnan päässä keskiarvosta.



ESIMERKKI 3

Marsu on Etelä-Amerikasta kotoisin oleva jyrsijä, josta on tullut suosittu lemmikkieläin. Marsun elinikä vaihtelee neljän ja kymmenen vuoden välillä. Marsujen elinikää tutkittiin tilastoimalla erään populaation eliniät.

Ikä	f
4	12
5	18
6	22
7	26
8	11
9	13
10	6

a) Millä välillä on niiden marsujen elinikä, jotka ovat enintään yhden keskihajonnan päässä keskiarvosta?
b) Maailman vanhin marsu eli 14-vuotiaaksi. Poikkeaako marsun ikä merkitsevästi aineiston keskiarvosta?

RATKAISU

a) Lasketaan ensin aineiston keskiarvo.

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{n} = \frac{12 \cdot 4 + 18 \cdot 5 + 22 \cdot 6 + 26 \cdot 7 + 11 \cdot 8 + 13 \cdot 9 + 6 \cdot 10}{108}$$
$$= 6{,}638...$$

Lasketaan keskihajonta laskimen tilastotoiminnolla.
Aineiston keskihajonta on $s = 1{,}6881...$

Yhden keskihajonnan päässä keskiarvosta olevien marsujen elinikä on vähintään $6{,}638... - 1{,}6881... \approx 4{,}9$ vuotta ja enintään $6{,}638... + 1{,}6881... \approx 8{,}3$ vuotta.

b) Lasketaan, kuinka vanha marsun tulisi olla, jotta marsun ikä olisi korkeintaan kahden keskihajonnan päässä keskiarvosta.

$$\bar{x} + 2s = 6{,}638... + 2 \cdot 1{,}6881... = 10{,}01...$$

Koska $10{,}01... < 14$, niin marsun ikä on yli kahden keskihajonnan päässä keskiarvosta. Marsun ikä poikkeaa siis merkitsevästi keskiarvosta.

VASTAUS

a) välillä 4,9–8,3 b) poikkeaa

Tehtäviä

E1 63. Liikennevirasto ja kaupunkien selvitysten mukaan maanteiden liikenteen yli 50 dB:n yömelulle altistuu noin 175 000 asukasta. Moottoriliikennetie kulkee pikkukaupungin halki. Yöliikenteen melua mitattiin viitenä arkiyönä ja saatiin seuraavat tulokset:

Yö	ma	ti	ke	to	pe
Melu (dB)	62	44	41	53	72

Laske taulukoimalla viiden yön melun keskihajonta.

64. Suomen elokuvateattereissa esitettyjen uutuuselokuvien määrät vuosilta 1991–2000 olivat seuraavat:

Vuosi	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Elokuvia	179	150	169	163	147	166	154	152	192	176

Lähde: Suomen elokuvasäätiö

Laske aineiston avulla,
a) kuinka monta uutuuselokuvaa keskimäärin esitettiin
b) mikä oli keskihajonta kyseisinä vuosina.

E2 65. Mäkihyppyvalmentaja on koonnut kilpailuja varten joukkuetta. Hänen täytyy valita vielä yksi jäsen sekä pienen mäen kilpailuun että joukkuemäen kilpailuun. Valmentaja katsoo Matin ja Tonin kuluneen kauden hyppytuloksia pienessä mäessä:

Matti	82	85	78	90	72	80	91	77	73	85	89	79
Toni	79	81	81	82	78	81	78	82	83	77	77	78

a) Kumpi hyppää keskimäärin pidempiä hyppyjä?
b) Kumpi hyppää tasaisemmin?

66. Suomen valtioiden 5 vuoden ja 10 vuoden obligaatioiden korot vaihtelivat vuosina 1995–2001 seuraavan tilaston mukaisesti.

Vuosi	1995	1996	1997	1998	1999	2000	2001
5 vuotta	7,93	6,03	4,86	4,3	4,07	5,27	4,54
10 vuotta	8,79	7,08	5,95	4,78	4,74	5,49	5,04

Lähde: Suomen pankki

Laske molemmille obligaatioille keskiarvo ja keskihajonta. Arvioi näiden avulla kumpaakin obligaatiotyyppiä.

67. Taulukkoon on luokiteltu tärkeimmät toimivat tulivuoret korkeuden perusteella. Laske aineiston avulla tulivuoren keskikorkeus. Kuinka suuri on korkeuksien keskihajonta?

Korkeus (m)	*f*
0–999	6
1000–1999	12
2000–2999	9
3000–3999	17
4000–4999	7
5000–5999	6
6000–6999	2

Lähde: Maailmantilasto 1998

68. Abiturientit saivat helmikuussa 1991 pitkässä ranskassa seuraavat määrät kouluarvosanoja:

Arvosana	4	5	6	7	8	9	10	yhteensä 152
Lukumäärä	0	8	10	30	36	39	29	

Piirrä pylväsdiagrammi aineistosta sekä laske taulukossa olevien arvosanojen keskiarvo ja keskihajonta. [K1992, 4b]

69. Ilmatieteen laitos määrittää vuoden kaikille kuukausille keskilämpötilan. Taulukossa on esitetty heinäkuun keskilämpötilat seitsemältä vuodelta Helsingissä ja Sodankylässä. Vertaile paikkakuntien keskilämpötiloja sopivia sijainti- ja hajontalukuja käyttäen.

	1995	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**
Helsinki (°C)	17	15	19,2	16,4	18,7	17	20,2	19,1
Sodankylä (°C)	14,1	12,9	15,9	15,5	14,7	15,5	14,9	15,6

Lähde: Ilmatieteen laitos

E3 70. Erään formulakuljettajan ajat sekunteina 21 kierrokselta olivat seuraavat:

142	148	131	135	120	125	138
135	128	129	151	145	142	138
150	152	155	128	130	132	133

a) Määritä aineiston keskihajonta.
b) Poikkeavatko nopein ja hitain ajoista merkitsevästi keskiarvosta?

71. Taulukossa on tietoja maakotkan poikasmääristä vuosina 1990–2001.

Vuosi	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001
Poikasia	87	53	90	72	89	92	123	104	119	114	118	127

Kuvaa jakaumaa pylväsdiagrammilla. Laske jakauman keskiarvo ja keskihajonta. Havaintoarvon sanotaan poikkeavan merkitsevästi keskiarvosta, jos poikkeama on vähintään kahden hajonnan suuruinen. Tutki, poikkeavatko maakotkan vuosien 1991 ja 2001 poikasmäärät merkitsevästi vuotuisen poikasmäärän keskiarvosta annetulla aikavälillä. [S2003, 3]

72. Avomaan kurkkujen massoja tutkittiin ottamalla 16 kurkun näyte-erä. Tulokset (grammoina) on esitetty taulukossa.

280	350	390	205	285	360	405	270
225	395	320	400	345	295	370	300

a) Luokittele massat tasavälisiin luokkiin siten, että pienin luokka on 200–249 g.
b) Laske aineiston keskiarvo ja keskihajonta.
c) Kasvimaalta löytyy 520 g painava jättikurkku. Poikkeaako kurkun massa merkitsevästi aineiston keskiarvosta?

Sekalaisia tehtäviä

73. Suomen sisäsaaristossa tehtiin keskikesällä mittauksia tuulen voimakkuudesta eri päivinä kolmen viikon ajan. Mittaustulokset (m/s) olivat seuraavat:

1. viikko	3,4	12,1	7,2	5,5	1,8	1,9	2,1
2. viikko	2,7	2,4	9,9	12,6	5,9	1,5	0,4
3. viikko	0,8	1,9	1,8	2,3	1,7	1,1	1,9

Määritä tuulen voimakkuuden keskihajonta kyseisenä ajanjaksona.

74. Johanna ja Elina osallistuivat taitoluistelun piirimestaruuskilpailuihin. Kuusi kilpailutuomaria arvioivat heidän ohjelmansa seuraavin pistein:
Johanna: 5,8; 5,7; 5,5; 5,8; 5,6 ja 5,7
Elina: 5,4; 5,9; 5,6; 5,9; 5,9 ja 5,6
Lopullisen sijoituksen määrää keskiarvo.
a) Kumman loppusijoitus on parempi?
b) Kumman pisteiden keskihajonta on suurempi?

75. Eräässä kokeessa annettiin suoritusten arvosanoiksi 0, 1, 2, 3, 4, 5 tai 6. Näiden prosenttiosuudet olivat seuraavat:

Arvosana	0	1	2	3	4	5	6
Osuus	5,80	10,99	17,54	24,78	19,95	15,48	5,46

Laske kokeen keskiarvo ja keskihajonta. [K2011, 7]

76. Vuoden 2011 asuntomessut järjestettiin Kokkolassa. Näyttelykohteina oli 39 erityyppistä ja erikokoista asuntoa kerrostaloasunnosta huvilaan. Asuntojen koko vaihteli noin 40 m^2 kaksioista 230 m^2 huvilaan. Asuntojen alat on koottu taulukkoon.

Ala (m^2)	**Asuntoja**
40–79	6
80–119	6
120–159	9
160–199	10
200–239	8

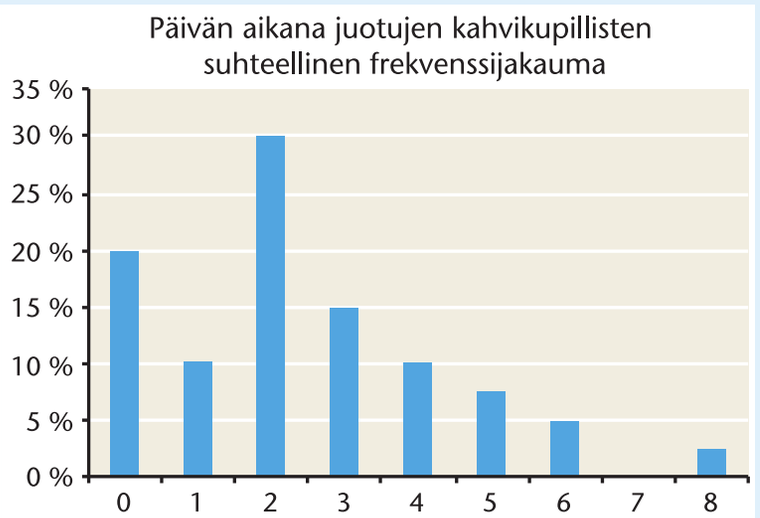
a) Määritä aineiston keskiarvo ja keskihajonta.
b) ”Rantakivi” oli yksi messujen suurimmista omakotitaloista. Sen pinta-ala oli 225 m^2. Onko talo messualueen rakennukseksi poikkeuksellisen suuri eli poikkeaako talon ala merkitsevästi messualueen talojen keskikoosta?

77. Lukujen x, $x + 4$, $3x - 2$ ja $4x$ keskiarvo on 16,25. Laske lukujen keskihajonta.

1.5 Normaalijakauma

POHDINTA 1

Oppilaskunta järjesti koulussa kyselyn siitä, kuinka monta kuppia kahvia opiskelijat juovat koulupäivän aikana. Alla on tutkimustuloksista muodostetun suhteellisen frekvenssijakauman kuvaaja.



Vastaa kuvaajan perusteella seuraaviin kysymyksiin.

Kuinka monta prosenttia koulun oppilaista juo kahvia

a) keskimäärin yhden kupillisen koulupäivässä

b) keskimäärin 2 kupillista koulupäivässä

c) korkeintaan 2 kupillista koulupäivässä

d) yli 2 kupillista koulupäivässä

e) 2–4 kupillista koulupäivässä

f) alle 4 kupillista koulupäivässä

g) vähintään 4 kupillista koulupäivässä?

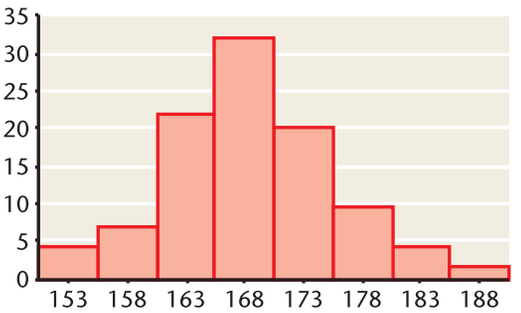
h) Jos tutkimukseen osallistui 154 koulun opiskelijaa, niin kuinka moni koulun oppilaista joi kahvia 2 kupillista koulupäivän aikana?

i) Millä todennäköisyydellä satunnaisesti valittu koulun oppilas joisi kahvia koulupäivän aikana 2–4 kupillista?

Normitus

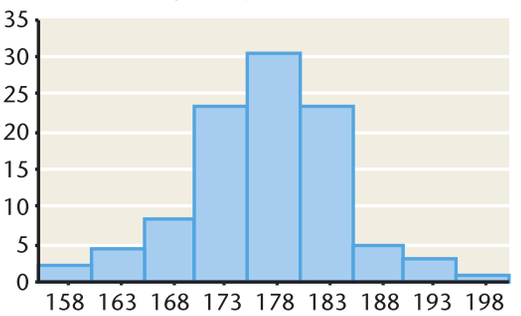
Tarkastellaan kahta sukupuolen mukaan kerättyä otosta ($n = 100$). Otoksista on muodostettu pituusjakaumat sukupuolen mukaan. Aineistojen keskiarvot ja keskihajonnat on esitetty kuvaajien alapuolella.

Naisten pituusjakauma ($n = 100$)



Keskiarvo $\bar{x} = 169$ cm
Keskihajonta $s \approx 7{,}3$ cm

Miesten pituusjakauma ($n = 100$)



Keskiarvo $\bar{x} = 177$ cm
Keskihajonta $s \approx 7{,}2$ cm

Naisten pituuksien keskiarvo on miesten keskiarvoa pienempi. Kummassakin jakaumassa keskihajonta on suurin piirtein yhtä suuri. Koska keskiarvot pituusjakaumissa poikkeavat toisistaan, ei muuttujan arvojen vertailu jakaumien kesken ole yleensä mielekästä.

Tarkastellaan kummassakin aineistossa henkilöä, jonka pituus on 180 cm.

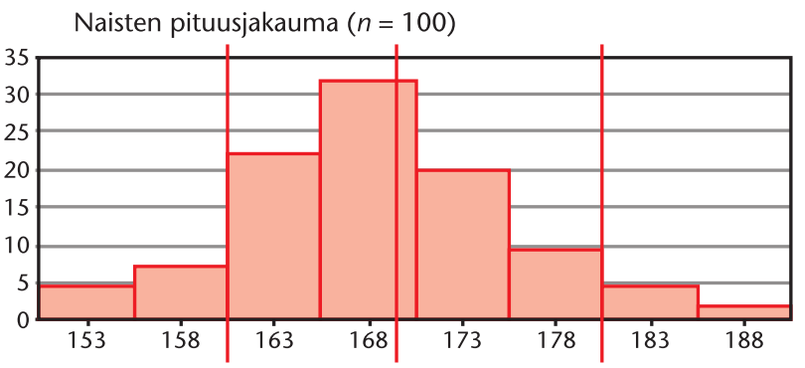
- Nainen, jonka pituus on 180 cm, on 11 cm:n eli $\dfrac{11\,\text{cm}}{7{,}3\,\text{cm}} \approx 1{,}5$ keskihajonnan päässä keskiarvosta.

- Mies, jonka pituus on 180 cm, on 3 cm:n eli $\dfrac{3\,\text{cm}}{7{,}2\,\text{cm}} \approx 0{,}4$ keskihajonnan päässä keskiarvosta.

Näitä muuttujan arvon etäisyyksiä keskiarvosta kutsutaan **normitetuiksi arvoiksi**. Miehen pituuden normitettu arvo (0,4) on pienempi kuin samanpituisen naisen vastaava normitettu arvo (1,5). Voidaan siis sanoa, että muihin samaa sukupuolta oleviin henkilöihin verrattuna 180 cm:n mittainen mies on lyhyempi kuin samanpituinen nainen.

Jos tarkastellaan muuttujan arvoa, joka on keskiarvoa pienempi, on normitettu arvo tällöin negatiivinen. Esimerkiksi 160 cm pitkä nainen on 9 cm keskipituutta lyhyempi. Tätä pituutta vastaava normitettu arvo on

$$\frac{-9\text{ cm}}{7{,}3\text{ cm}} \approx -1{,}2$$



160 cm sijaitsee noin 1,2 keskihajonnan päässä keskiarvosta. Normitettu arvo on –1,2.

Keskiarvo $\bar{x} = 169$ cm
Keskihajonta $s \approx 7{,}3$ cm

180 cm sijaitsee noin 1,5 keskihajonnan päässä keskiarvosta. Normitettu arvo on 1,5.

Muuttujan normitettua arvoa merkitään yleensä kirjaimella z. Normitettujen arvojen avulla voidaan vertailla tilanteita, joissa keskiarvot ja keskihajonnat poikkeavat toisistaan.

Normitettu arvo

Muuttujan arvoa x vastaava normitettu arvo on

$$z = \frac{x - \bar{x}}{s}$$

- $\bar{x}$ on keskiarvo ja
- s on keskihajonta.

ESIMERKKI 1 Afrikkalaisen metsämuurahaisen keskipituus on 1,2 cm ja pituuksien keskihajonta 0,2 cm. Kuningatar voi kasvaa jopa 2,0 cm:n pituiseksi. Jos ihmisrodun hallitsijat kasvaisivat vastaavalla tavalla, kuinka pitkä Tarja Halonen olisi? Suomalaisnaisten keskipituus on 169 cm ja pituuden keskihajonta 7,3 cm.

RATKAISU Jos muurahaiskuningatar ja Tarja Halonen olisivat suhteessa muihin lajitovereihinsa yhtä pitkiä, pituuksien normitettujen arvojen täytyisi olla yhtä suuret.

Normitetaan ensin muurahaiskuningattaren pituus.

$$z_{\text{muurahainen}} = \frac{2{,}0\text{ cm} - 1{,}2\text{ cm}}{0{,}2\text{ cm}} = 4$$

Merkitään Tarja Halosen pituutta kirjaimella x. Tällöin

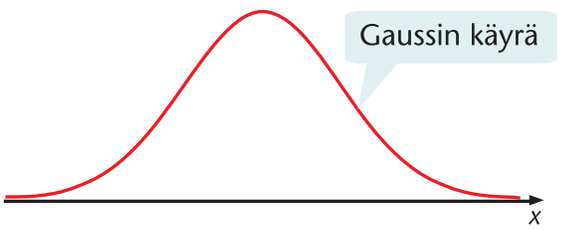
$$\begin{aligned} \frac{x-169}{7{,}3} &= 4 \qquad \big| \cdot 7{,}3 \\ x - 169 &= 29{,}2 \\ x &= 198{,}2 \\ x &\approx 198\,(\text{cm}) \end{aligned}$$

VASTAUS Tarja Halonen olisi 198 cm pitkä.

Normitettu normaalijakauma

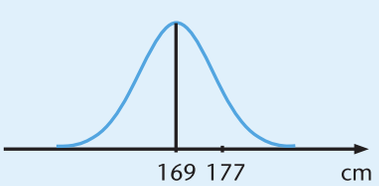
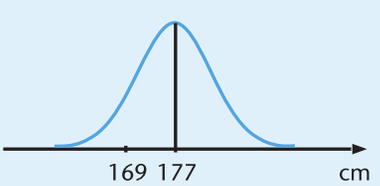
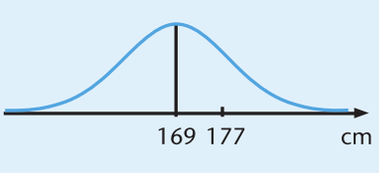
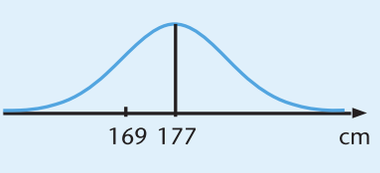
Jatkuvat muuttujat, esimerkiksi pituus ja paino, voivat saada vaihtelualueellaan kaikki mahdolliset reaalilukuarvot. Jatkuvan muuttujan jakaumaa kutsutaan lyhyesti **jatkuvaksi jakaumaksi.**

Jatkuvista jakaumista tärkein ja tunnetuin on **normaalijakauma**. Jakauman muotoa kuvaa Gaussin käyrä, joka on nimetty keksijänsä *Carl Friedrich Gaussin* mukaan.



Normaalijakauman muoto riippuu sen keskiarvosta ja keskihajonnasta.

- Jakauma on symmetrinen keskiarvon suhteen.
- Jakauman huippukohta on keskiarvon kohdalla, joten jakauman sijainti vaaka-akselilla riippuu keskiarvosta.
- Keskihajonta kuvaa jakauman leveyttä: mitä suurempi hajonta, sitä leveämpi käyrä.

	Keskiarvo 169 cm	Keskiarvo 177 cm
Keskihajonta 7 cm		
Keskihajonta 15 cm		

Normaalijakauman avulla kuvataan yleensä muuttujan arvojen suhteellisia frekvenssejä. Koska suhteellinen frekvenssi kuvaa myös tilastollista todennäköisyyttä, normaalijakaumaa voidaan käyttää myös todennäköisyyslaskennassa. Todennäköisyyslaskennassa keskiarvon käsitettä vastaa **odotusarvo** μ. Tämä nimitys on vakiintunut myös normaalijakauman kohdalla käytettäväksi keskiarvon sijaan. Keskihajonnan kohdalla käytetään yleensä kreikkalaista kirjainta σ.

Normaalijakauma

Kun muuttujan x jakauma on Gaussin käyrän muotoinen, sanotaan muuttujan noudattavan normaalijakaumaa N:

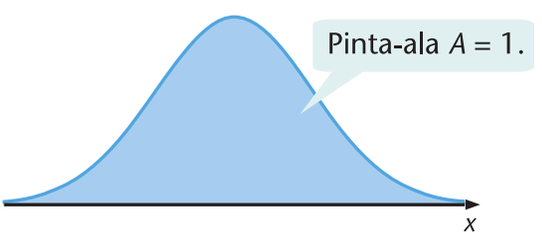
$$x \sim N(\mu,\sigma)$$

- x on muuttuja
- μ on odotusarvo (keskiarvo)
- σ on keskihajonta

Luetaan: ”x noudattaa normaalijakaumaa, jonka odotusarvo on μ ja keskihajonta σ”.

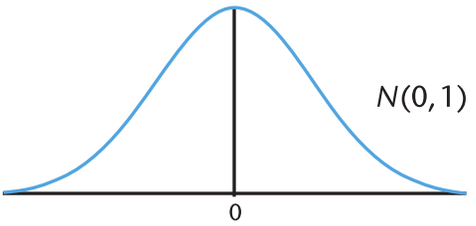
Kun suhteellisia frekvenssejä lasketaan yhteen, saadaan suhteelliset summafrekvenssit, jotka kuvaavat kertymää tiettyyn muuttujan arvoon mennessä. Jatkuvan muuttujan jakauman kohdalla tämä summa lasketaan käyrän ja x-akselin väliin jäävän pinta-alan avulla. Tätä samaa ideaa sovelletaan pohdintatehtävässä 1.

Kun suhteelliset frekvenssit lasketaan yhteen, niiden summa on aina 100 %. Tästä syystä Gaussin käyrän ja x-akselin väliin jäävä pinta-ala on 1 = 100 %.



Gaussin käyrän alle jäävän pinta-alan tarkat arvot saadaan **kertymäfunktion** Φ (”fii”) avulla. Kertymäfunktio kertoo tiettyyn muuttujan arvoon mennessä kertyneen pinta-alan suuruuden.

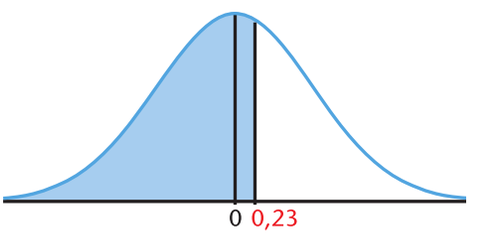
Normitettu normaalijakauma on jakauma, jonka odotusarvo on 0 ja keskihajonta 1. Normitettu normaalijakauma on siis $N(0, 1)$.



Normitetun normaalijakauman kertymäfunktion arvoja on koottu taulukkokirjan taulukkoon. Taulukon avulla voidaan määrittää Gaussin käyrän alle jäävien pinta-alojen tarkkoja arvoja.

Esimerkiksi väritetty pinta-ala saadaan kertymäfunktion avulla.

$\Phi(0{,}23)$



Luetaan kertymäfunktion arvo taulukosta. Kaikki kertymäfunktion arvot ovat välillä 0–1. Tästä syystä taulukossa olevat luvut ovat kertymäfunktion arvon neljä desimaalia.

2. Tarkasteltavan arvon toinen desimaali haetaan taulukon ensimmäiseltä riviltä.

x	0	1	2	3	…
0,0	0,5000	5040	5080	5120	…
0,1	5398	5438	5478	5517	…
0,2	5793	5832	5871	5910	…
…	…	…	…	…	…

1. Tarkasteltavan arvon kokonaisosa ja ensimmäinen desimaali haetaan taulukon ensimmäisestä sarakkeesta.

$\Phi(0{,}23) = 0{,}5910$

Tämän perusteella

- aineistossa on 59,10 % arvoja, jotka ovat pienempiä kuin 0,23
- todennäköisyys, että muuttujan arvo on pienempi kuin 0,23 on 59,10 %.

Prosenttiosuuksia on taulukoitu vain normitetulle normaalijakaumalle. Muiden normaalijakaumien muuttujien arvot on normitettava ennen taulukon käyttöä.

ESIMERKKI 2 Erään heinäkasvin pituus noudattaa likimain normaalijakaumaa. Keskipituus on 13 cm ja keskihajonta on $\sigma = 2{,}4$ cm. Kuinka monen prosentin todennäköisyydellä satunnaisesti valitun heinän pituus on alle 15 cm?

RATKAISU Keskipituus on sama kuin odotusarvo eli $\mu = 13$ (cm).

Normitetaan muuttujan arvo on $x = 15$ (cm).

Normitus:
$z = \frac{x - \bar{x}}{s} = \frac{x - \mu}{\sigma}$

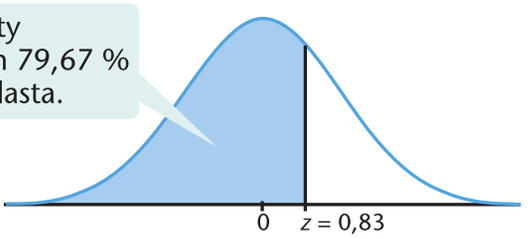
$$z_{15} = \frac{15 - 13}{2{,}4} = 0{,}8333\ldots \approx 0{,}83$$

Pyöristetään, koska taulukon normitetut arvot ovat kahden desimaalin tarkkuudella.

Luetaan taulukosta tätä normitettua arvoa vastaava prosenttiluku eli kertymäfunktion arvo

$$\Phi(0{,}83) = 0{,}7967 = 79{,}67\ \%$$

Väritetty alue on 79,67 % koko alasta.



Todennäköisyys, että heinäkasvin pituus on alle 15 cm, on $79{,}67\ \% \approx 80\ \%$.

VASTAUS Heinäkasvin pituus on alle 15 cm todennäköisyydellä 80 %.

ESIMERKKI 3 Kunta järjesti abiturienteille tasokokeen matematiikassa. Kokeeseen osallistui yhteensä 4536 abiturienttia. Kokeen tulokset noudattivat normaalijakaumaa $N(36; 8{,}6)$.

a) Millä todennäköisyydellä abiturientti saa kokeesta yli 40 pistettä?

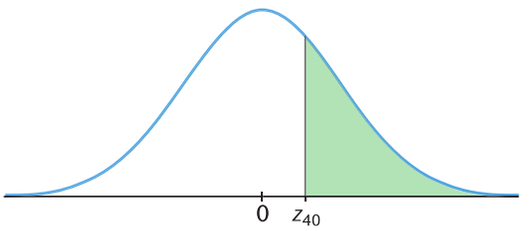
b) Kuinka moni kokeeseen osallistuneista sai yli 40 pistettä?

RATKAISU

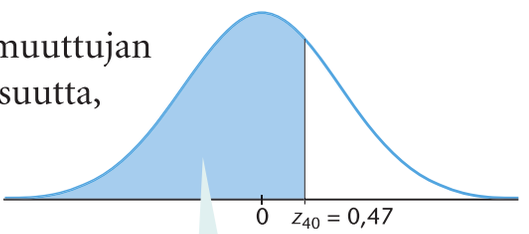
a) Kokeen tulosten odotusarvo on $\mu = 36$ ja keskihajonta $\sigma = 8{,}6$.

Normitetaan muuttujan arvo 40 (pistettä).

$$z_{40} = \frac{40 - 36}{8{,}6} = 0{,}4651... \approx 0{,}47$$



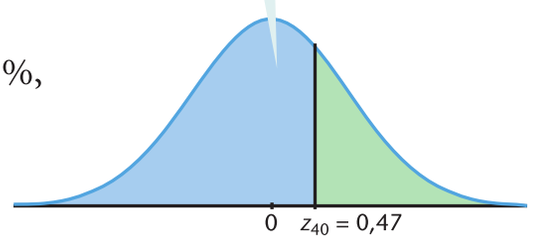
Kertymäfunktion arvo kuvaa muuttujan arvoon mennessä kertynyttä osuutta, joka tarkoittaa kuviossa rajan 0,47 vasemmalle puolelle jäävää pinta-alaa.



$\Phi(0{,}47) = 0{,}6808 = 68{,}08\ \%$

Pinta-ala on 68,08 %.

Koska koko pinta-ala on 100 %, saadaan kysytty pinta-ala vähennyslaskun avulla.



$$P(\text{abiturientti saa yli 40 pistettä}) = 100\ \% - 68{,}08\ \% = 31{,}92\ \% \approx 0{,}32$$

b) Laskettu todennäköisyys on sama kuin abiturienttien suhteellinen frekvenssi eli prosenttisuus.

Lasketaan kuinka paljon on 31,92 % 4536 abiturientista:

$$0{,}3192 \cdot 4536 = 1447{,}89 \approx 1448$$

VASTAUS

a) Todennäköisyys on 0,32.

b) Abiturientteja, jotka saivat yli 40 pistettä, oli 1448 kappaletta.

Normaalijakauman käyttöä

Normaalijakaumatehtävissä tilanteen hahmottamiseksi kannattaa piirtää mallikuva, jossa näkyy

- pinta-ala, joka halutaan ratkaista
- pinta-alaan liittyvät muuttujan arvot.

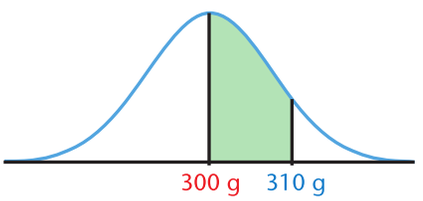
Normittamisen ja kertymäfunktion arvojen määrittämisen lisäksi tilanteiden ratkaisemiseksi hyödynnetään normaalijakauman symmetrisyyttä.

ESIMERKKI 4

Keksipaketin massa x noudattaa normaalijakaumaa $N(300\text{ g}, 7\text{ g})$. Laske todennäköisyys, että kaupan hyllyltä otetun keksipaketin massa on 300 g – 310 g.

RATKAISU

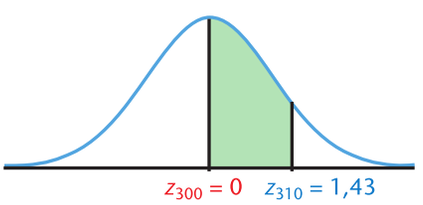
Keksipaketin massan odotusarvo on $\mu = 300$ g ja keskihajonta $\sigma = 7$ g. Piirretään tilanteesta mallikuva.



Normitetaan massojen arvot:

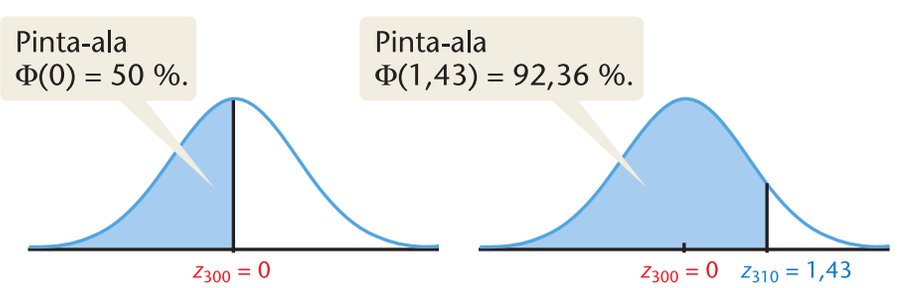
$$z_{300} = \frac{300-300}{7} = 0 \quad \text{ja} \quad z_{310} = \frac{310-300}{7} = 1{,}4285\ldots \approx 1{,}43$$

Sama tilanne kuvattuna normitettujen arvojen avulla:



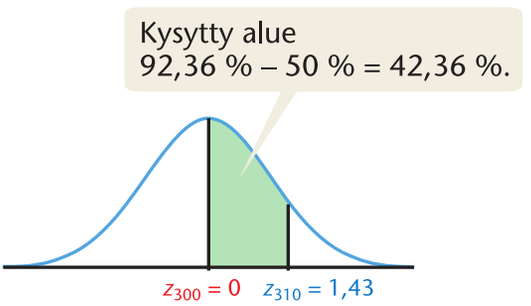
Luetaan taulukosta vastaavat kertymäfunktion arvot:

$$\Phi(0) = 0{,}5000 = 50\ \% \qquad \Phi(1{,}43) = 0{,}9236 = 92{,}36\ \%$$



Normitettujen arvojen 0 ja 1,43 väliin jäävä prosenttiosuus saadaan erotuksen avulla:

$$\Phi(1{,}43) - \Phi(0) = 92{,}36\ \% - 50\ \% = 42{,}36\ \%$$



Tällöin

$$P(\text{paketin massa on } 300\text{ g} - 310\text{ g}) = 0{,}4236 \approx 0{,}42$$

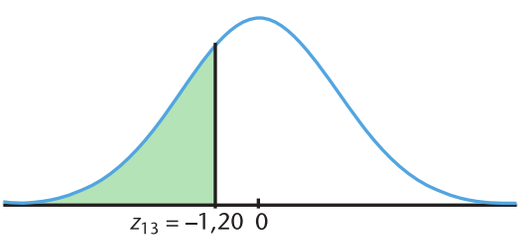
VASTAUS Keksipaketin massa on 300 g – 310 g todennäköisyydellä 0,42.

ESIMERKKI 5 Matematiikan kokeessa opiskelijoiden saamat pistemäärät x noudattivat normaalijakaumaa $N(19, 5)$. Millä todennäköisyydellä satunnaisesti valittu opiskelija sai korkeintaan 13 pistettä?

RATKAISU Kokeen odotusarvo on $\mu = 19$ (pistettä) ja keskihajonta $\sigma = 5$ (pistettä).

Tulosta 13 pistettä vastaava normitettu arvo on

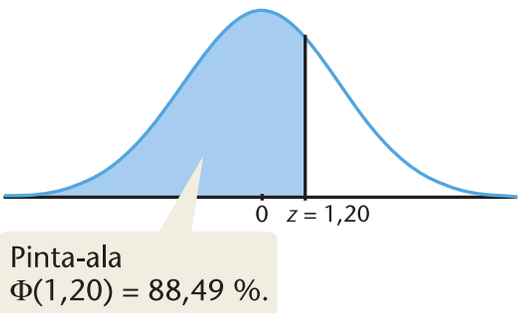
$$z_{13} = \frac{13-19}{5} = -1{,}20$$



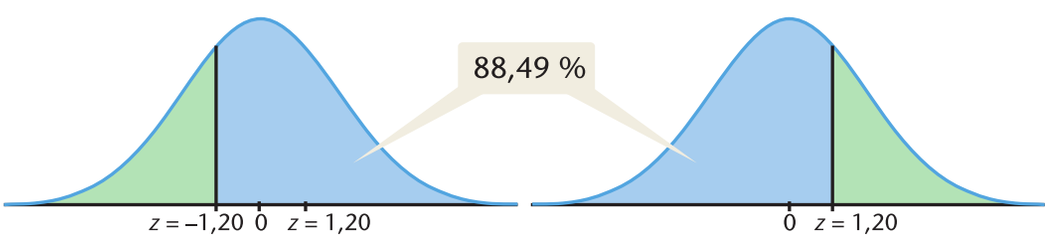
Koska $z_{13} = -1{,}20$ on negatiivinen, taulukossa ei ole sille kertymäfunktion arvoa.

Sen sijaan se on annettu symmetria-akselin oikealla puolella arvolle $z = 1{,}20$.

$$\Phi(1{,}20) = 88{,}49\ \%$$



Symmetriasta johtuen arvon $z = 1{,}20$ jälkeinen pinta-ala on yhtä suuri kuin kysytty pinta-ala.



Koska koko aluetta kuvaa 100 %, haluttu kertymäfunktion arvo saadaan vähennyslaskulla:

$$\begin{aligned}\Phi(-1{,}20) &= 100\ \% - \Phi(1{,}20)\\ &= 100\ \% - 88{,}49\ \%\\ &= 11{,}51\ \%\end{aligned}$$

Kysytty todennäköisyys on

$$P(\text{opiskelija sai korkeintaan 13 pistettä}) = 0{,}1151 \approx 0{,}12$$

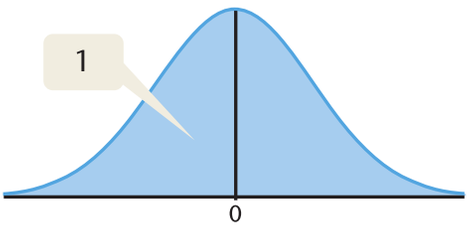
VASTAUS Todennäköisyys, että opiskelija sai korkeintaan 13 pistettä, oli 0,12.

Millä todennäköisyydellä muuttujan arvo on suurempi kuin 0,23?

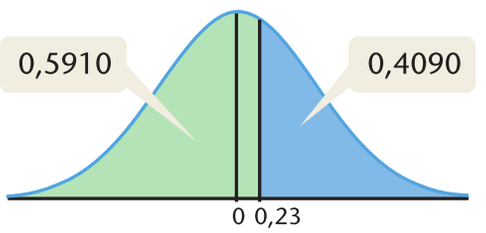
eli

Kuinka monta prosenttia muuttujan arvoista on suurempia kuin 0,23?

Normaalijakauman ja vaaka-akselin väliin jäävä pinta-ala on 1.



Todennäköisyys, että muuttujan arvo on suurempi kuin 0,23, on $1 - 0{,}5910 = 0{,}4090 = 40{,}90\ \%$.

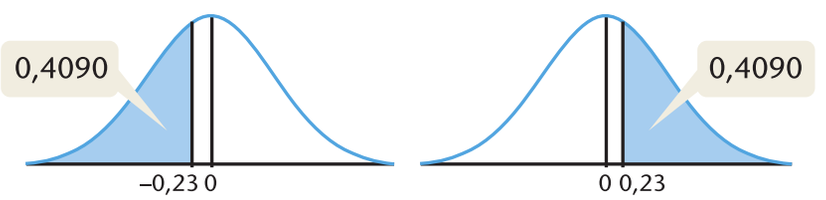


Millä todennäköisyydellä muuttujan arvo on pienempi kuin –0,23?

eli

Kuinka monta prosenttia muuttujan arvoista on pienempiä kuin –0,23?

Koska jakauma on symmetrinen, arvon –0,23 vasemmalle puolelle jäävä pinta-ala on yhtä suuri kuin arvon 0,23 oikealle puolelle jäävä pinta-ala.



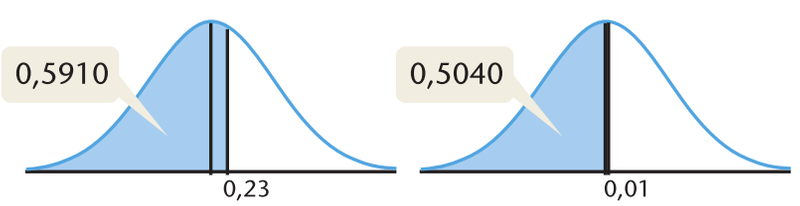
Todennäköisyys, että muuttujan arvo on pienempi kuin –0,23, on $0{,}4090 = 40{,}90\ \%$.

Millä todennäköisyydellä muuttujan arvo on välillä 0,01–0,23?

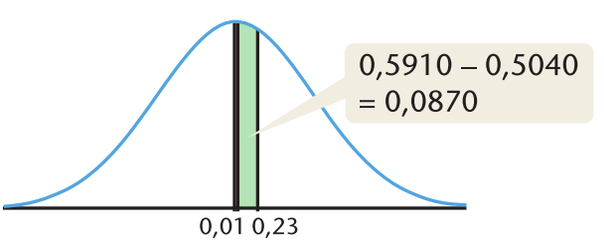
eli

Kuinka monta prosenttia muuttujan arvoista on välillä 0,01–0,23?

Selvitetään ensin erikseen, kuinka suuret ovat arvojen 0,23 ja 0,01 vasemmalle puolelle jäävät pinta-alat.



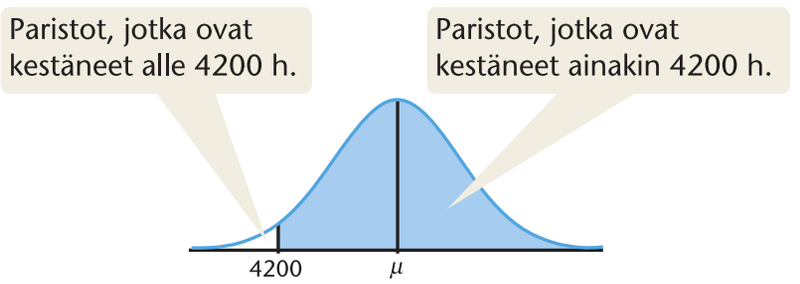
Näiden pinta-alojen erotuksena saadaan selville muuttujan arvojen väliin jäävä osuus.



Todennäköisyys, että muuttujan arvo on välillä 0,01–0,23, on $0{,}0870 = 8{,}70\ \%$.

ESIMERKKI 6 Oletetaan, että television kaukosäätimen pariston ikä noudattaa normaalijakaumaa keskihajonnan ollessa 160 tuntia. Määritä pariston iän odotusarvo, kun satunnaisesti valittu paristo kestää ainakin 4200 tuntia todennäköisyydellä 0,95.

RATKAISU Olkoon pariston keston odotusarvo μ.



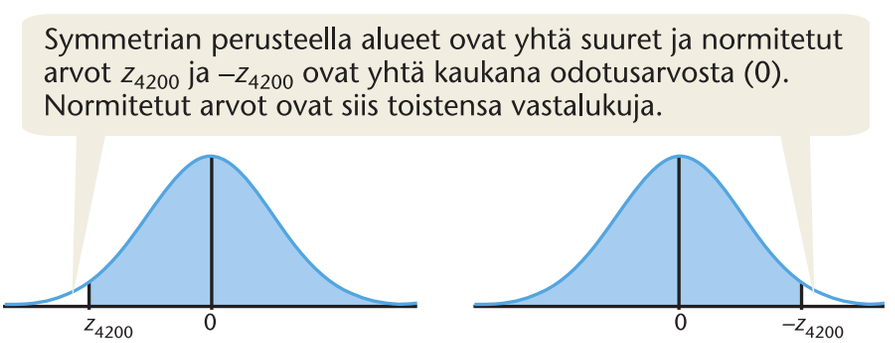
Arvoa 4200 h vastaa normitettu arvo

$$z_{4200} = \frac{4200 - \mu}{160}$$

Koska $P(\text{paristo kestää ainakin 4200 h}) = 0{,}95$,
$P(\text{paristo kestää alle 4200 h}) = 1 - 0{,}95 = 0{,}05$.

$\Phi(z_{4200}) = 0{,}05$

Normitettua arvoa z_{4200} vastaa siis todennäköisyys 0,05. Tätä arvoa ei löydy taulukosta, joten tarkastellaan tilannetta kuvaajien avulla.



Etsitään taulukosta arvo $(-z_{4200})$, jonka kertymäfunktion arvo on mahdollisimman lähellä lukua 0,95.

	0	1	2	3	4	5	6	7
1,6	9452	9463	9474	9484	9495	9505	9515	9525

94,95 % 95,05 %

Luvut 0,9495 ja 0,9505 ovat yhtä lähellä lukua 0,95. Valitaan lukuarvo, joka ylittää 95 %. Tätä vastaava normitettu arvo on 1,65 eli $-z_{4200} = 1{,}65$. Tällöin $z_{4200} = -1{,}65$.

$$\frac{4200-\mu}{160} = -1{,}65 \qquad \mid \cdot 160$$
$$4200-\mu = -264$$
$$-\mu = -4464 \qquad \mid \cdot(-1)$$
$$\mu = 4464$$

VASTAUS Odotusarvon on oltava 4464 tuntia.

Tehtäviä

Normitus ja normitettu normaalijakauma

78. Maikin kissan massa on 4,8 kg. Rodun keskimääräinen massa on 4,3 kg ja keskihajonta 0,7 kg. Vilmalla on taas koira, jonka massa on 12,8 kg. Koirarodun keskimääräinen massa on 10,2 kg ja keskihajonta 0,9 kg. Kumman lemmikin massa on suhteessa suurempi?

E1

79. Oletetaan, että naisen keskimääräinen elinikä on 82 vuotta ja iän keskihajonta 4,5 vuotta. Erään papukaijalajin ikä voi olla jopa 140 vuotta, vaikka keskimääräinen elinikä onkin 92 vuotta keskihajonnan ollessa 9,5 vuotta. Kuinka vanha ihmisen olisi oltava, jotta se eläisi suhteessa yhtä vanhaksi kuin 140-vuotias lintu?

80. Rasmus sai historian kokeesta 8+. Kokeen keskiarvo oli 7,2 ja keskihajonta 0,8. Ruotsin kokeessa Rasmus menestyi suhteessa yhtä hyvin, vaikka saikin arvosanaksi vain 7+ keskihajonnan ollessa sama. Mikä oli ruotsin kokeen keskiarvo?

81. Normita muuttujan 35 arvo seuraavissa normaalijakaumissa.
a) $N(35, 4)$ b) $N(39, 4)$ c) $N(31, 4)$

82. Olkoon muuttujan x odotusarvo 0 ja keskihajonta 1. Oletetaan, että muuttuja x noudattaa normaalijakaumaa. Laske taulukkoa käyttäen todennäköisyydet
a) $P(x < 0{,}25)$ b) $P(x < 1{,}37)$ c) $P(x > 1{,}37)$

E2

83. Erään valtakunnallisen matematiikan kokeen pistemäärät noudattivat normaalijakaumaa. Kokeen pisteiden keskiarvo oli 56 pistettä ja keskihajonta 18 pistettä. Kuinka monen prosentin todennäköisyydellä satunnaisesti valitussa vastauspaperissa tehtävien yhteispistemäärä on alle 60 pistettä?

84. Satunnaismuuttuja x noudattaa normaalijakaumaa $N(20, 4)$. Laske todennäköisyys, että muuttujan arvo on
a) enintään 25 b) vähintään 25.

85. Hylkeen massa noudattaa normaalijakaumaa, kun odotusarvo on 32 kg ja keskihajonta 1,7 kg. Laske todennäköisyys, että eläintieteilijän tutkittavana olevan hylkeen massa on
a) korkeintaan 34 kg b) ainakin 34 kg.

E3 86. Erääseen kokeeseen osallistui 233 oppilasta. Tulokset noudattivat normaalijakaumaa $N(29; 1{,}967)$.
a) Laske todennäköisyys, että oppilas saa yli 33 pistettä.
b) Kuinka moni oppilas sai yli 33 pistettä?

87. Tulitikkurasiassa olevien tikkujen määrä noudattaa normaalijakaumaa $N(45, 4)$. Millä todennäköisyydellä satunnaisesti valitussa rasiassa on
a) alle 50 tikkua b) yli 52 tikkua?

Normaalijakauman käyttöä

88. Eräällä metsäpalstalla kasvavien puiden keskipituus oli 15,0 m ja keskihajonta 3,5 m.
a) Inka valitsi kaksi puuta, joiden pituudet olivat 16,7 m ja 20,5 m. Normita puiden pituudet.
b) Millä todennäköisyydellä metsästä satunnaisesti valittu puu oli pituudeltaan yli 20,5 metriä?
c) Millä todennäköisyydellä metsästä satunnaisesti valittu puu oli pituudeltaan 16,7–20,5 metriä?

E4 89. Luukaksen koulumatkaan käyttämä aika noudattaa normaalijakaumaa. Luukaksen koulumatka kestää keskimäärin 35 minuuttia ja koulumatkan keskihajonta on 7 minuuttia. Millä todennäköisyydellä Luukaksen koulumatkan kesto on 30–40 minuuttia?

90. Erään älykkyysosamäärätestin tulokset noudattavat normaalijakaumaa odotusarvona 100 ja keskihajontana 15. Kuinka monella prosentilla testiin osallistuvista henkilöistä älykkyysosamäärä on
a) alle 120 b) yli 120
c) 90–120 d) alle 90?

E5 91. Karkkipussin massa noudattaa normaalijakaumaa $N(85\text{ g}, 4\text{ g})$. Karkkipussien valmistuslinjalta valitaan satunnaisesti yksi pussi. Millä todennäköisyydellä karkkipussin massa on alle 83 g?

92. Myyrän massa noudattaa normaalijakaumaa $N(99\text{ g}, 8\text{ g})$. Millä todennäköisyydellä myyräpopulaatiosta satunnaisesti valitun myyrän massa on
a) yli 90 g b) alle 90 g?

93. Sähkölampun kestoikä noudattaa normaalijakaumaa $N(5000\text{ h}, 300\text{ h})$. Millä todennäköisyydellä satunnaisesti valittu sähkölamppu kestää
a) vähintään 5500 tuntia b) 5200–5500 tuntia?

94. Erään kissarodun hännän pituus noudattaa normaalijakaumaa keskiarvon ollessa 17,5 cm. Keskihajonta on 2,5 cm. Millä todennäköisyydellä kissan hännän pituus on välillä [15 cm, 20 cm]?

E6 95. Rocktapahtumassa myydään lakritsinauhoja, joiden pituuden keskihajonta on 4,0 cm. Pituus noudattaa likimain normaalijakaumaa. Laske lakritsinauhojen odotusarvo, kun enintään 0,9 m pituisia lakritsinauhoja on kaikista nauhoista
a) 50 % b) 85 %.

96. Metsäntutkimuslaitoksen koealalla mitattiin 120 männyn läpimitta rinnan korkeudelta 0,1 cm:n tarkkuudella. Tulokset jakautuivat likimain normaalisti. Mäntyjä, joiden läpimitta oli yli 26 cm, oli 20 %. Laske mäntyjen läpimitan odotusarvo, kun keskihajonta on 2,65 cm.

97. Hammaslääkärikäynti kestää keskimäärin 20 min. Oletetaan, että kestoaika on normaalisti jakautunut $N(20, \sigma)$. Mikä keskihajonnan σ on oltava, ettei käynti 98 % varmuudella ylitä 35 minuuttia?

98. Erään yrityksen valmistamien akselien läpimitta on normaalisti jakautunut keskiarvon ollessa 150,00 mm. Kuinka suuri saa läpimitan hajonta enintään olla, jotta läpimitta ylittäisi hylkäämisrajan 150,20 mm korkeintaan 0,5 % todennäköisyydellä? [S1992, 8]

99. Oletetaan, että matematiikan ylioppilaskokeen tulokset (jotka ovat välillä 0,…, 60 pistettä) jakautuvat likimain normaalisti keskiarvona 30 ja keskihajontana 10. Määritä tämän perusteella laudaturin pisteraja, kun tavoitteena on antaa laudatureja enintään viidelle prosentille osallistujista. [S2011, 10]

Sekalaisia tehtäviä

100. Erään havupuulajin keskipituus on 18 m ja keskihajonta 1,4 m. Puistossa elävän puuvanhuksen pituus on jo 20,2 m. Vastaavasti erään heinän keskipituus on 13 cm ja keskihajonta on 2,3 cm. Kuinka pitkä heinän olisi oltava, jotta se olisi suhteessa yhtä pitkä kuin mainittu puiston puu?

101. Olkoon muuttujan x odotusarvo 0 ja keskihajonta 1. Oletetaan, että muuttuja x noudattaa normaalijakaumaa. Laske taulukkoa käyttäen todennäköisyydet
a) $P(x < 0{,}45)$ b) $P(x < -2)$ c) $P(-1 < x < 0{,}98)$.

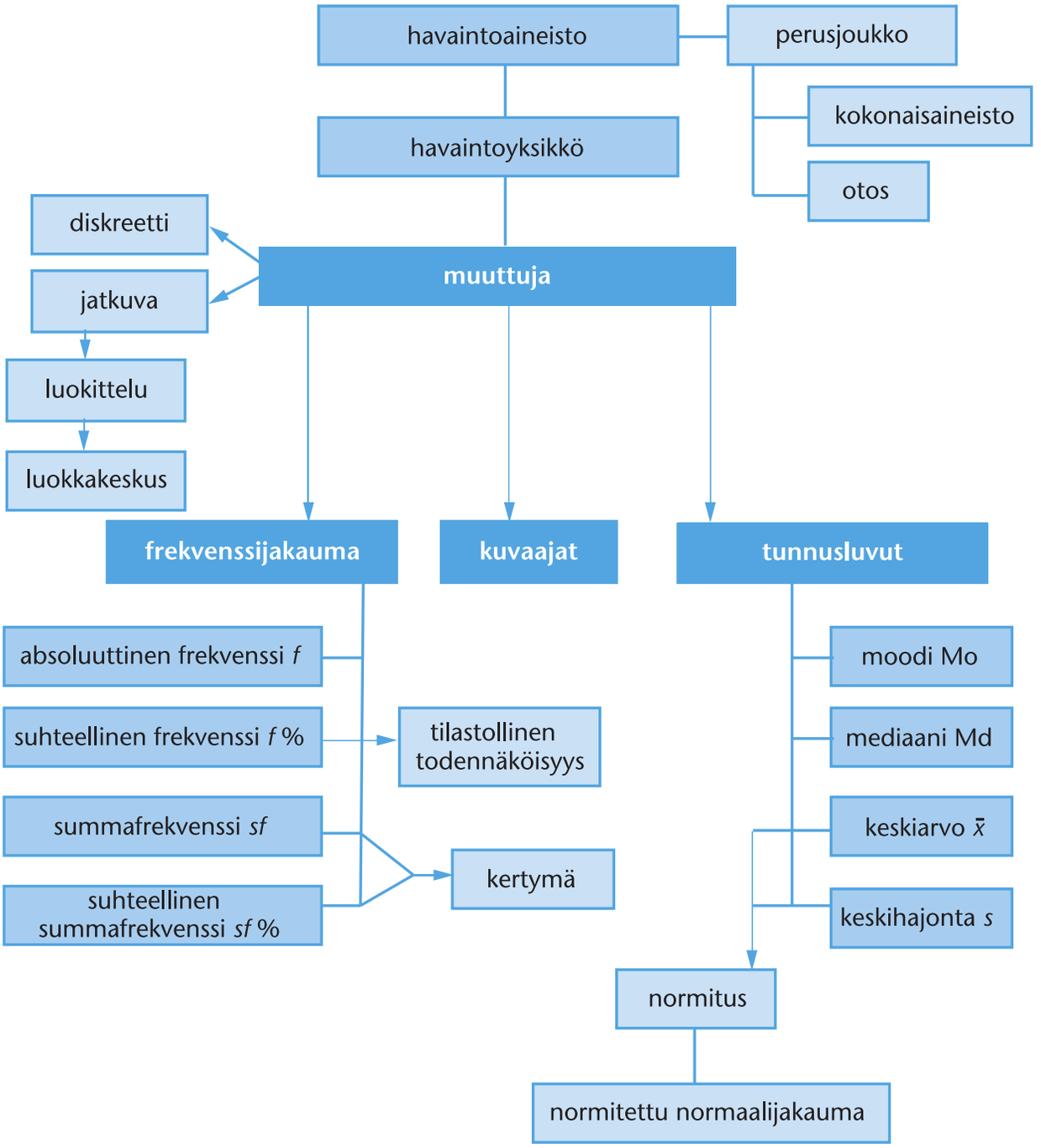
102. Muuttuja x noudattaa normaalijakaumaa $N(45, 12)$. Millä todennäköisyydellä
a) $x > 50$ b) $x > 40$ c) $x < 40$?

103. Matematiikan kesäkurssille osallistui 60 oppilasta. Lopputentin pisteet noudattivat normaalijakaumaa $N(25; 3{,}5)$. Millä todennäköisyydellä satunnaisesti valittava kurssilainen sai
a) yli 30 pistettä b) alle 20 pistettä c) 20–30 pistettä?

104. Hernekeittopurkin massa noudattaa normaalijakaumaa odotusarvon ollessa 405 g. Kuinka suuri keskihajonta on, kun 15 % todennäköisyydellä satunnaisen purkin massa on yli 410 g?

105. Eräässä lyhyen matematiikan ylioppilaskokeessa keskiarvo oli 27,36 ja keskihajonta 12,23 pistettä. Jos oletetaan pistemäärän noudattavan normaalijakaumaa, niin
a) mikä on alin pistemäärä, jolla saa laudaturin, jos laudaturarvosanojen osuudeksi halutaan enintään 5 % kaikista kokelaista
b) kuinka monta prosenttia kokelaista hyväksytään, jos alin hyväksytty pistemäärä on 12? [S2005, 14]

Keskeisiä käsitteitä



2 Todennäköisyys

2.1 Malleja todennäköisyyden määrittämiseksi

POHDINTA 1 Miten valitsisit seuraavissa tilanteissa? Perustele.

a) Laatikossa A on 60 hedelmää, joista 15 on pilaantuneita. Laatikossa B on 80 hedelmää, joista 18 on pilaantuneita. Kummasta laatikosta ottaisit hedelmän, jos joutuisit nostamaan sen sokkona?

b) Musta hevonen on osallistunut 27 kilpailuun, joista se on voittanut 12. Valkoinen hevonen on osallistunut 35 kilpailuun, joista se on voittanut 15. Kumman hevosen puolesta löisit vetoa?

c) Kolikonheitossa olet saanut 12 peräkkäisellä heitolla kruunan. Kannattaako lyödä vetoa sen puolesta, että seuraavalla heitolla saat klaavan?

d) Kesäjuhlilla heitetään pyöreitä muovilautasia neljän metrin päässä oleviin maahan piirrettyihin ympyröihin. Pientä lautasta (halkaisija 12 cm) heitetään ympyrään, jonka halkaisija on 0,8 m. Suurempaa lautasta (halkaisija 18 cm) heitetään ympyrään, jonka halkaisija on 1,0 m. Lautasen on pudottava kokonaan ympyrän sisälle. Kummassa heitossa on helpompi onnistua?

e) Markkinoilla järjestetään kahdet lukuarpajaiset. Arpajaisissa A voit saada minkä tahansa luvun väliltä 1–2. Voitat, jos saamasi luku on suurempi kuin 1,5. Arpajaisissa B voit saada minkä tahansa luvun välillä 1–4. Voitat, jos saamasi luku on suurempi kuin 3. Kumpiin arpajaisiin kannattaa osallistua?

Todennäköisyyslaskenta eli **stokastiikka** sai alkunsa uhkapeleihin liittyvistä ongelmista. Jo 1500-luvulla matemaatikot laativat rahapelejä ja vedonlyöntisuhteita kuvaavia alkeellisia malleja.

Todennäköisyyslaskennan isänä voidaan pitää ranskalaista matemaatikkoa *Blaise Pascalia* (1623–1662). Hän muotoili ensimmäiset onnistuneet todennäköisyyslaskennan laskusäännöt *Chevalier de Mérénä* tunnetun pelurin ehdotuksesta.

Pascal oli rajoittunut ongelmiin, joissa tulosmahdollisuudet olivat symmetrisiä, kuten kolikonheitossa. *Jakob Bernoulli* yleisti Pascalin teorian myös epäsymmetrisiin tapauksiin. Näin tuli määriteltyä **yleinen binomijakauma**.

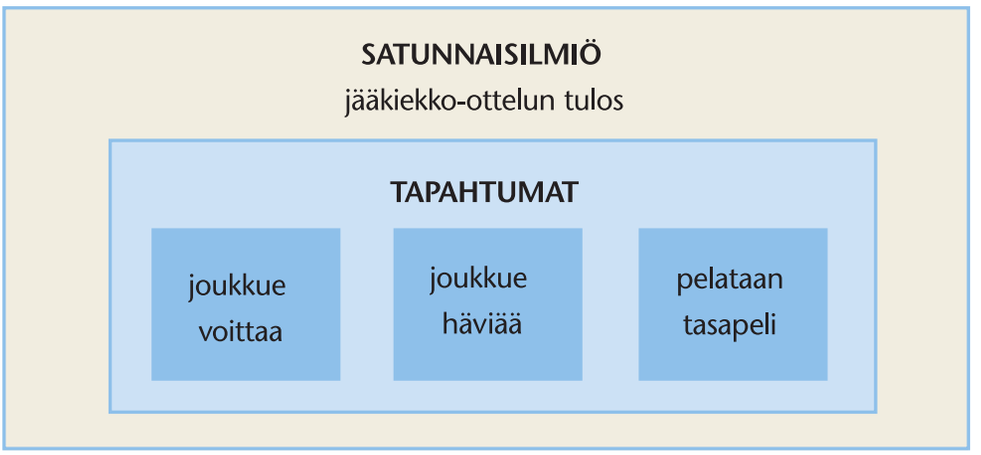
Nykyään stokastiikka on yksi matematiikan tärkeimmistä osaalueista, ja sitä sovelletaan lukuisilla tieteenaloilla. Biologiassa lasketaan populaatioiden säilymistä ja ominaisuuksien periytymisen todennäköisyyttä. Fysiikassa ja kemiassa kuvataan erilaisilla todennäköisyysfunktioilla elektronin sijaintia atomissa ja ennustetaan radioaktiivisen laskeuman vaikutuksia. Meteorologit käyttävät hyväkseen todennäköisyyslaskentaa ennustaessaan säätilaa.

Todennäköisyys liittyy moniin arkipäivän ilmiöihin. Voidaan käyttää esimerkiksi seuraavia ilmauksia:

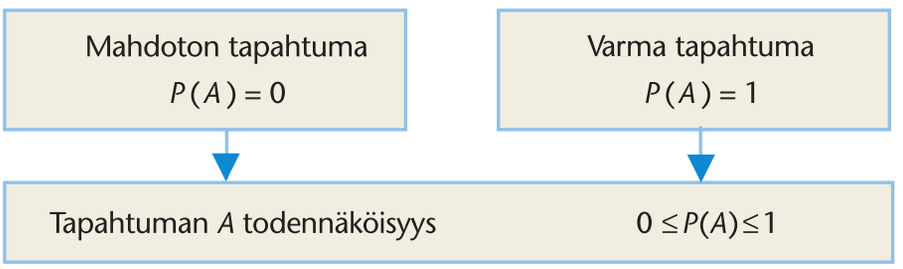
- Todennäköisesti myöhästyn tapaamisesta.
- On hyvin todennäköistä, että huomenna sataa.
- Todennäköisesti Suomi voittaa Ruotsin jääkiekko-ottelussa.

Yhteistä edellisille esimerkeille on se, ettei tulevista tapahtumista ole täyttä varmuutta. Jos sattumalla on osuutensa tilanteessa, ilmiötä kutsutaan **satunnaisilmiöksi**. Edellä esitellyt ilmaukset sisältävät satunnaisilmiöt: ”tapaaminen”, ”sää” ja ”jääkiekko-ottelun tulos”.

Satunnaisilmiöt sisältävät eri **tapahtumia**. Esimerkiksi satunnaisilmiö ”jääkiekko-ottelun tulos” sisältää tapahtumat: ”joukkue voittaa”, ”joukkue häviää” ja ”pelataan tasapeli”.

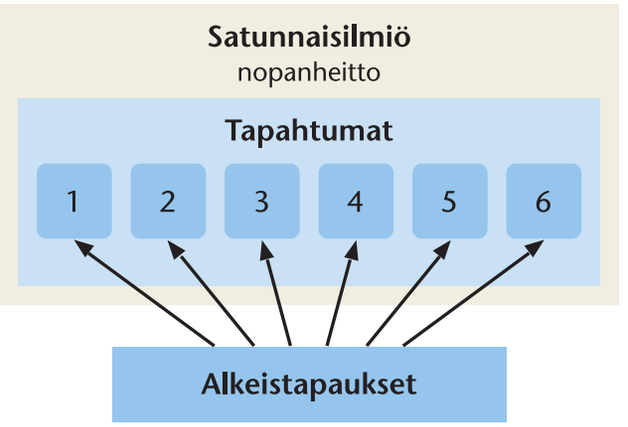


Tapahtuman A todennäköisyydestä käytetään merkintää $P(A)$. P tulee englannin kielen sanasta *probability*. Luku $P(A)$ ilmaisee tapahtuman A epävarmuuden astetta. Sen arvo on nollan ja ykkösen välillä.



Klassinen todennäköisyys tarkastelee todennäköisyyksiä **symmetristen tapausten** avulla, jolloin jokainen tapaus on yhtä mahdollinen. Todellisuudessa täysin symmetrisiä tapauksia on harvoin, mutta esimerkiksi nopanheittoa voidaan tarkastella klassisen todennäköisyyden avulla.

Yhden nopan heitossa voidaan saada tulokseksi mikä tahansa silmäluvuista 1, 2, 3, 4, 5, ja 6. Nämä kuusi eri heittotulosta muodostavat **alkeistapausten** joukon. Alkeistapauksia ovat kaikki mahdolliset tapahtumat, joita tilanteessa voi sattua.



Suotuisia tapauksia ovat ne alkeistapaukset, joiden todennäköisyyttä ollaan laskemassa eli ne tapaukset, jotka toteuttavat tapahtuman. Tapahtuman A klassinen todennäköisyys $P(A)$ määritellään suotuisien tapausten ja alkeistapausten lukumäärien suhteena.

Jos esimerkiksi halutaan laskea todennäköisyyttä tapahtumalle ”nopanheitossa saadaan silmäluku 5”, suotuisia tapauksia on vain yksi.

$$P(\text{nopanheitossa saadaan silmäluvuksi 5}) = \frac{1}{6}$$

Tapahtuman A todennäköisyys

$$P(A) = \frac{\text{suotuisien tapausten lukumäärä}}{\text{kaikkien alkeistapausten lukumäärä}}$$

ESIMERKKI 1

Vuonna 2011 valitussa eduskunnassa on 85 naiskansanedustajaa ja 115 mieskansanedustajaa. Naisvaltaisin puolue on SDP, jonka 42 edustajasta 27 on naisia. Valitaan arpomalla yksi edustaja. Millä todennäköisyydellä valittu henkilö on

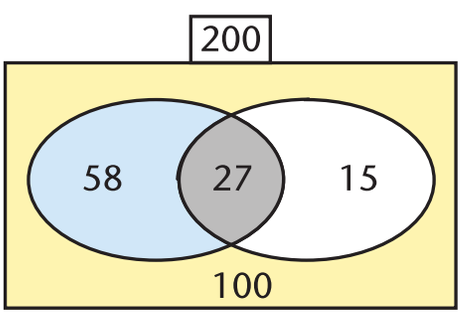
a) nainen, mutta ei ole sosiaalidemokraatti

b) mies, mutta ei ole sosiaalidemokraatti?

RATKAISU

Käytetään ratkaisemisessa apuna Venn-diagrammia. Kaksi soikeaa aluetta kuvaavat naiskansanedustajia ja SDP:n kansanedustajia.

- Suorakulmio kuvaa koko kansanedustajien joukkoa:
  $85 + 115 = 200$
- Alueiden leikkaus eli harmaa alue kuvaa SDP:n naiskansanedustajia, joita on 27.
- Sininen alue kuvaa naiskansanedustajia, jotka eivät ole SDP:n jäseniä:
  $85 - 27 = 58$
- Valkoinen alue kuvaa SDP:n mieskansanedustajia:
  $42 - 27 = 15$
- Keltainen alue kuvaa mieskansanedustajia, jotka eivät ole SDP:n jäseniä:
  $200 - (58 + 27 + 15) = 100$



a) P(henkilö on nainen, mutta ei ole sosiaalidemokraatti)

$$= \frac{58}{200} = 0{,}29$$

b) P(henkilö on mies, mutta ei ole sosiaalidemokraatti)

$$= \frac{100}{200} = 0{,}5$$

VASTAUS a) 0,29 b) 0,5

ESIMERKKI 2

Illallisseurueessa on 18 henkilöä. Heistä 6 ottaa alkuruokaa, 10 ottaa jälkiruokaa ja 4 ottaa molempia. Millä todennäköisyydellä joukosta satunnaisesti valittu henkilö ei ota alku- eikä jälkiruokaa?

RATKAISU

Ratkaistaan tehtävä käyttäen apuna taulukkoa. Sijoitetaan annetut tiedot ensin taulukkoon.

Tehtävän voi ratkaista myös Venn-diagrammin avulla kuten esimerkissä 1.

Henkilöitä, jotka

- syövät alkuruoan, mutta eivät jälkiruokaa, on $6 - 4 = \mathbf{2}$
- syövät jälkiruoan, mutta eivät alkuruokaa, on $10 - 4 = \mathbf{6}$
- eivät syö alkuruokaa, on $18 - 6 = \mathbf{12}$
- eivät syö jälkiruokaa, on $18 - 10 = \mathbf{8}$.

	Jälkiruokia syövät henkilöt	Henkilöt, jotka eivät syö jälkiruokaa	Yhteensä
Alkuruokia syövät henkilöt	4	2	6
Henkilöt, jotka eivät syö alkuruokaa	6	?	12
Yhteensä	10	8	18

Henkilöitä, jotka eivät syö alku- eivätkä jälkiruokaa, on $8 - 2 = 6$.

$$P(\text{valittu henkilö ei ota alku- eikä jälkiruokaa}) = \frac{6}{18} = \frac{1}{3}$$

VASTAUS

$\frac{1}{3} \approx 0{,}33$

ESIMERKKI 3

Aku, Iines, Roope, Touho ja Hannu seisovat taksijonossa. Paikalle saapuu taksi, johon mahtuu vain kolme matkustajaa. He päättävät arpoa keskuudestaan kolme onnekasta, jotka pääsevät taksiin. Millä todennäköisyydellä

a) Hannu pääsee taksiin
b) Aku ja Iines pääsevät taksiin
c) Roope tai Touho pääsee taksiin, mutta eivät molemmat?

RATKAISU

Taulukoidaan ensin kaikki mahdolliset arvonnan tulokset. Merkitään kutakin henkilöä alkukirjaimen avulla:

Aku A Iines I Roope R Touho T Hannu H

Taksiin voi päästä yksi seuraavista kolmen hengen ryhmistä:

AIR	AIT	AIH
ART	ARH	ATH
IRT	IRH	ITH
RTH		

Erilaisia tulosmahdollisuuksia on siis 10.

a) Tapahtumia, joissa Hannu pääsee taksiin, on 6 (AIH, ARH, ATH, IRH, ITH, RTH). Tällöin

$$P(\text{Hannu pääsee taksiin}) = \frac{6}{10} = \frac{2}{5} = 0{,}6$$

b) Aku ja Iines pääsevät taksiin tapahtumissa AIR, AIT ja AIH eli niitä on kolme. Tällöin

$$P(\text{Aku ja Iines pääsevät taksiin}) = \frac{3}{10} = 0{,}3$$

c) Tapahtumia, joissa taksiin pääsee Roope tai Touho, on 6 (AIR, AIT, ARH, ATH, IRH, ITH). Tällöin

$$P(\text{Roope tai Touho pääsee taksiin}) = \frac{6}{10} = \frac{2}{5} = 0{,}6$$

VASTAUS

a) 0,6 b) 0,3 c) 0,6

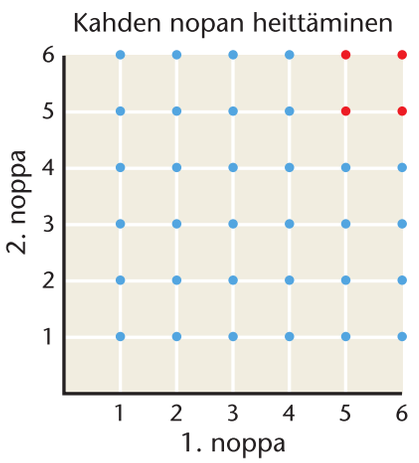
ESIMERKKI 4 Heitetään noppaa kaksi kertaa. Millä todennäköisyydellä molemmilla kerroilla saadaan silmäluvuksi vähintään 5?

RATKAISU Muodostetaan ensin kaikki mahdolliset tapaukset. Merkitään heittoja lukupareilla. Esimerkiksi merkinnällä (4, 3) tarkoitetaan tilannetta, jossa ensimmäisellä heitolla saadaan 4 ja toisella heitolla 3. On huomattava, että heittotulokset (4, 3) ja (3, 4) kuvaavat siis **eri tilanteita**.

Tapa 1 Havainnollistetaan alkeistapauksia taulukon avulla.

(1,1)	(2,1)	(3,1)	(4,1)	(5,1)	(6,1)
(1,2)	(2,2)	(3,2)	(4,2)	(5,2)	(6,2)
(1,3)	(2,3)	(3,3)	(4,3)	(5,3)	(6,3)
(1,4)	(2,4)	(3,4)	(4,4)	(5,4)	(6,4)
(1,5)	(2,5)	(3,5)	(4,5)	(5,5)	(6,5)
(1,6)	(2,6)	(3,6)	(4,6)	(5,6)	(6,6)

Tapa 2 Havainnollistetaan alkeistapauksia koordinaatistossa.



Näiden 36 pisteparin joukko kuvaa kahden heiton alkeistapauksia. Taulukkoon ja koordinaatistoon on merkitty punaisella suotuisat tapaukset eli ne heitot, joissa molempien heittojen silmäluku on vähintään 5. Niitä on neljä kappaletta. Tällöin

$$P(\text{molempien heittojen silmäluku vähintään 5}) = \frac{4}{36} = \frac{1}{9} \approx 0{,}11$$

VASTAUS Kysytty todennäköisyys on $\frac{1}{9} \approx 0{,}11$.

Yksinkertaisissa tilanteissa todennäköisyys voidaan määrittää etsimällä kaikki mahdolliset alkeistapaukset ja poimimalla niiden joukosta suotuisat tapaukset. Tämä onnistuu, jos alkeistapausten joukko on pieni ja niillä on sama todennäköisyys. Alkeistapausten etsimisessä kannattaa käyttää erilaisia apukeinoja, kuten taulukoita ja kaavioita.

Jos alkeistapauksia on ääretön määrä, alkeistapausten joukkoa voidaan kuvata geometrisella suureella. Suotuisien tapausten joukko on tällöin tietty osa tästä suureesta. Jos erimerkiksi alkeistapauksia esittää tietyn mittainen jana, suotuisien tapausten joukko on osa tästä janasta.

Tapahtuman *A* geometrinen todennäköisyys

$$P(A) = \frac{\text{suotuisan osuuden suuruus}}{\text{koko geometrisen kuvion suuruus}}$$

Tavallisimmat geometrian suureet, joita todennäköisyyslaskennassa käytetään, ovat

- pituus
- pinta-ala
- tilavuus
- kulma.

Geometrista todennäköisyyttä sovelletaan varsinkin jatkuvan muuttujan tapauksissa. Esimerkki jatkuvasta muuttujasta on aika.

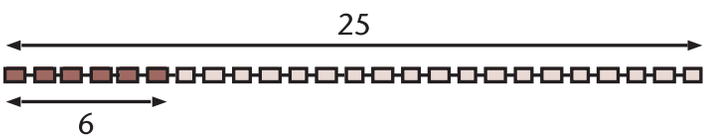
ESIMERKKI 5 Lähtöasemalle saapuu 25 minuutin välein paikallisjuna, joka odottaa matkustajia 6 minuuttia. Liisa ei tiedä junan aikataulua saapuessaan asemalle. Millä todennäköisyydellä hän

a) pääsee heti junaan

b) joutuu odottamaan junaa yli 10 minuuttia?

RATKAISU Junia saapuu lähtöasemalle 25 minuutin jaksoissa. Jokaisella tällaisella aikavälillä on ääretön määrä ajanhetkiä. Jokaisen aikavälin ja ajanhetken luetteloiminen ei ole mahdollista, joten alkeistapaukset tutkitaan geometrisesti.

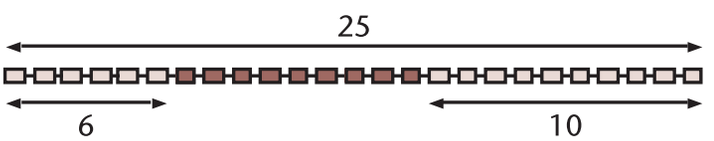
Tarkastellaan jotakin yksittäistä 25 minuutin jaksoa. Havainnollistetaan sitä geometrisesti janalla, jonka pituus on 25 yksikköä. Tähän janaan sisältyvä 6 yksikön pituinen osuus kuvaa junan seisomista lähtöasemalla.



a) Liisa pääsee junaan heti, jos hän tulee asemalle niiden 6 minuutin aikana, jotka juna seisoo lähtöasemalla.

$$P(\text{pääsee heti junaan}) = \frac{6}{25} = 0{,}24 \approx 0{,}2$$

b) Liisa joutuu odottamaan junaa yli 10 minuuttia, jos juna ei ole asemalla ja seuraavan junan saapumiseen on yli 10 minuuttia.



$$P(\text{joutuu odottamaan yli 10 min}) = \frac{25-6-10}{25}$$
$$= \frac{9}{25} = 0{,}36 \approx 0{,}4$$

VASTAUS a) 0,2 b) 0,4

ESIMERKKI 6

Roosa ja Liisa sopivat tapaavansa kahvilassa klo 14 ja 15 välillä. Kumpikin saapuu satunnaisena hetkenä kyseisen tunnin aikana. Millä todennäköisyydellä

a) molemmat saapuvat ennen klo 14.20

b) toinen joutuu odottamaan toisen saapumista yli 30 min?

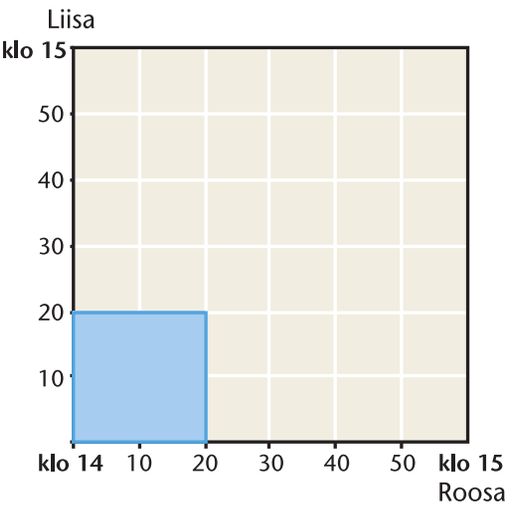
RATKAISU

a) Ajanhetket, jolloin Roosa ja Liisa saapuvat, voidaan havainnollistaa geometrisesti neliönä. Sivun pituus muodostuu klo 14 ja 15 välissä olevista ajanhetkistä.

Kaikkia alkeistapauksia kuvaa neliö, jonka ala on $A = 60^2 = 3600$.

Suotuisia tapauksia kuvaa kuvioon muodostuvan neliön ala

$A = 20 \cdot 20 = 400$

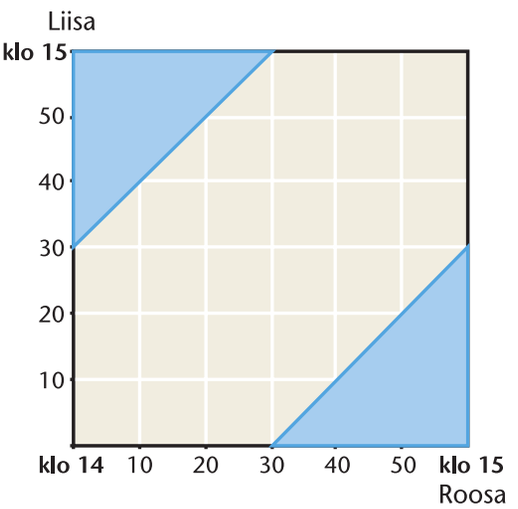


$$P(\text{molemmat saapuvat ennen klo 14.20}) = \frac{400}{3600} = \frac{2}{18} = \frac{1}{9}$$

b) Suotuisia tapauksia kuvaa kaksi kuvioon muodostunutta yhtä suurta kolmiota.

$$A_{\text{kolmiot}} = 2 \cdot \frac{1}{2} \cdot 30 \cdot 30 = 900$$

$$P(\text{joutuu odottamaan yli 30 min}) = \frac{900}{3600} = \frac{1}{4}$$



VASTAUS a) $\frac{1}{9}$ b) $\frac{1}{4}$

Tehtäviä

E1 106. Luokkaretkelle osallistuu 28 lukiolaista. Heistä 12 haluaa käydä taidemuseossa ja 13 eläintarhassa. Viisi opiskelijaa haluaa käydä kummassakin. Laadi Venn-diagrammi tilanteesta. Millä todennäköisyydellä ryhmästä satunnaisesti valittu opiskelija
a) käy vain toisessa paikassa
b) ei käy kummassakaan paikassa?

107. Myymälässä on 5500 erilaista tuotetta, joista luomutuotteita on 120. Erilaisia vihanneksia on myytävänä 50, joista luomuvihannesten osuus on 15 tuotetta. Valitaan kaupasta satunnaisesti yksi myytävä tuote. Millä todennäköisyydellä valittu tuote
a) on luomutuote, mutta ei ole vihannes
b) ei ole luomutuote, mutta on vihannes
c) ei ole luomutuote eikä ole vihannes?

E2 108. Erään yhdistyksen naisjäsenistä 17 käy työssä ja 19 jäsenellä on lapsia. Työssä käyviä äitejä on 14. Naisista kuusi ei käy työssä eikä ole äiti. Laadi tietojen avulla taulukko. Millä todennäköisyydellä satunnaisesti valittu naisjäsen
a) ei ole äiti
b) ei käy työssä?

109. Ryhmän 30 henkilöstä 17 pelaa tennistä ja 10 jääpalloa. Kuusi harrastaa molempia lajeja. Millä todennäköisyydellä tästä ryhmästä satunnaisesti valittu henkilö
a) ei harrasta kumpaakaan lajia
b) harrastaa vain toista lajia?

E3 110. Linnan ruokavarastossa on yksi ankka ja yksi hanhi. Lisäksi sieltä löytyy porsaanpaisti, naudanpaisti ja jänis. Kokki valmistaa kahden liharuokalajin illallista. Hän päättää arpoa tarjottavat lihat. Millä todennäköisyydellä illallisella
a) tarjotaan jänistä
b) tarjotaan lintua
c) ei tarjota porsasta?

111. Heitetään kolmea kolikkoa. Millä todennäköisyydellä saadaan
a) kolme klaavaa
b) ainakin kaksi klaavaa?

112. Koulun matematiikkakilpailun karsinnoissa viisi opiskelijaa sai täsmälleen saman pistemäärän. Tupu, Hupu ja Lupu olivat näiden viiden parhaan joukossa. Koulu saa lähettää loppukilpailuun kolme edustajaa. Heidät valitaan arvalla. Millä todennäköisyydellä kaikki ankanpojat pääsevät loppukilpailuun?

E4 113. Heitetään kahta noppaa. Millä todennäköisyydellä
a) molemmilla nopilla saadaan sama silmäluku
b) silmälukujen summa on vähintään 9
c) ainakin toinen silmäluvuista on vähintään 5?

114. Mikä on todennäköisin pistesumma kahden nopan heitossa? [S1999, 4a]

E5 115. Lossimatka kestää 20 minuuttia yhteen suuntaan ja lossi seisoo laiturissa 5 minuuttia. Millä todennäköisyydellä saapuva auto pääsee suoraan lossiin?

116. Tikkataulussa on yhdeksän rengasta, joiden leveys on 2 cm. Keskellä taulua on ympyrä, jonka säde on 1 cm ja pistearvo 10. Jokke heittää umpimähkään yhden tikan tauluun. Millä todennäköisyydellä hän
a) saa pisteluvuksi 10
b) saa pisteluvuksi vähintään 8
c) heittää pistelukujen 4 ja 5 väliselle kehälle?

117. Kattila, jonka pohjan halkaisija on 15 cm ja tilavuus 4,5 litraa, on täynnä puuroa. Puurossa on manteli. Joulupöydässä puuroa kuluu 8 cm:n kerros. Millä todennäköisyydellä manteli jää kattilaan?

118. Kivi heitetään ympyrään umpimähkään. Millä todennäköisyydellä se osuu lähemmäksi ympyrän keskipistettä kuin sen kehää? [K1998, 8a]

119. Puu, joka on muodoltaan likimäärin kärjellään seisova suora ympyräkartio ($h = 19$ m ja $r = 3{,}2$ m), kaatuu mielivaltaiseen suuntaan. Mikä on todennäköisyys, että se osuu puun tyvestä 15 m:n päässä seisovaan henkilöön? Henkilön mittoja ei tarvitse ottaa huomioon. [K1996, 9a]

E6 120. Otto ja Eeli tapaavat ravintolassa klo 18 ja 19 välillä. Kumpikin saapuu paikalle satunnaisena hetkenä kyseisen tunnin aikana. Millä todennäköisyydellä

a) Eeli saapuu ennen klo 18.20 ja Otto vasta 18.30 jälkeen
b) molemmat saapuvat ennen klo 18.30
c) Otto joutuu odottamaan Eeliä yli 10 minuuttia?

121. Tietokone arpoo kaksi reaalilukua 0:n ja 10:n väliltä. Millä todennäköisyydellä molemmat arvotut luvut ovat suurempia kuin 8?

122. Neliön kärkipisteet ovat (0, 0), (2, 0), (0, 2) ja (2, 2). Piste (x, y) valitaan umpimähkään neliön sisältä. Millä todennäköisyydellä

a) x-koordinaatti on pienempi kuin 1
b) y-koordinaatti on suurempi kuin 1,75
c) x- ja y-koordinaatit ovat suurempia kuin 1,8?

Sekalaisia tehtäviä

123. Taulukossa on tutkimustulokset 500 pariston kestosta. Millä todennäköisyydellä

a) satunnainen paristo kestää alle 200 tuntia
b) jo 100 tuntia kestänyt paristo kestää vielä 100 tuntia?

Kesto (h)	Kpl
110–99	8
100–199	245
200–299	235
300–399	12

124. Maatilalla on 240 lintua kotieläimenä. Linnuista 200 on valkoisia. Kanojen osuus linnuista on 150. Valkoisia kanoja on 120. Illalliseksi tarvitaan yksi lintu. Emäntä valitsee linnun arpomalla. Millä todennäköisyydellä pataan päätyy
a) valkoinen lintu, joka ei kuitenkaan ole kana
b) kana, joka ei ole valkoinen
c) lintu, joka ei ole valkoinen eikä ole kana?

125. Kaupunkiin on saapunut kiertävä taidenäyttely. Esillä on yhteensä 112 maalausta. Ranskalaisia ei-impressionistisiä maalauksia on yhtä monta kuin yhteensä näyttelyssä on impressionistisiä maalauksia. Impressionistisistä maalauksista neljäsosa on ranskalaisia. Maalauksia, jotka eivät ole impressionistisiä eivätkä ranskalaisia, on yhteensä 80. Taideopiskelijan tehtävänä on maalata kopio yhdestä teoksesta. Kopioitava teos valitaan arvalla. Millä todennäköisyydellä hän joutuu maalaamaan
a) impressionistisen maalauksen
b) ranskalaisen maalauksen
c) ranskalaisen impressionistisen maalauksen?

126. Dodekaedrin muotoisessa nopassa on 12 tahkoa, jotka on numeroitu silmäluvuilla yhdestä kahteentoista. Heitetään kahta noppaa. Millä todennäköisyydellä
a) molemmilla saadaan 10
b) molemmilla saadaan korkeintaan 5?

127. Eduskunnan paternosterhissi eli ”herrahissi” muodostuu avointen osastojen ketjusta, joka kulkee pysähtymättä ylös ja alas. Hissin saapuessa kaksi SDP:n ja kolme Kokoomuksen kansanedustajaa odottaa pääsyä hissiin. Hissiin mahtuu kerralla vain kaksi ihmistä. He arpovat kaksi onnekasta. Millä todennäköisyydellä hissiin
a) pääsee kaksi Kokoomuksen kansanedustajaa
b) pääsevät molemmat SDP:n edustajat?

128. Erään koulun kaikki opiskelijat ovat iältään vähintään 15-vuotiaita, mutta kukaan ei ole vielä täyttänyt 20 vuotta. Valitaan umpimähkään kaksi opiskelijaa. Millä todennäköisyydellä
a) molemmat ovat täyttäneet 18 vuotta
b) toinen on jo täyttänyt 18 vuotta, mutta toinen ei?

2.2 Komplementti ja kertolaskusääntö

POHDINTA 1

Seuraavissa lauseissa on kuvattu tapahtumia, jotka sisältävät yhden tai useamman mahdollisen tilanteen. Kirjoita jokaisesta tapahtumasta ns. vastatapahtuma, joka sisältää ne tapaukset, jotka eivät sisälly annettuun tapahtumaan.

a) Huomenna sataa vettä.

b) Noppaa yhden kerran heitettäessä saadaan silmäluvuksi 5 tai 6.

c) Noppaa yhden kerran heitettäessä saadaan silmäluvuksi **korkeintaan** 3.

d) Viiden henkilön joukossa on **ainakin** yksi mies.

e) Kymmenen oppilaan ryhmässä on **korkeintaan** kaksi vasenkätistä.

f) Kolmen Ässä-arvan joukossa on kolme voittoarpaa.

POHDINTA 2

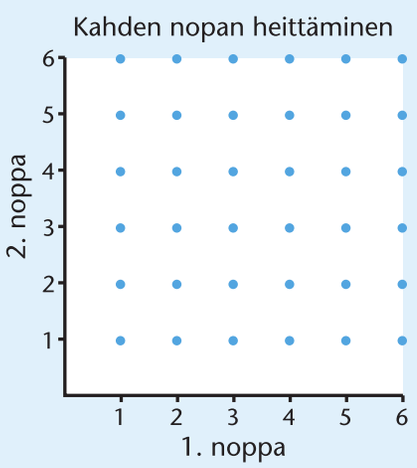
Noppaa heitetään kaksi kertaa.

a) Millä todennäköisyydellä ensimmäisellä heitolla saadaan silmäluku 3?

b) Millä todennäköisyydellä toisella heitolla saadaan silmäluku 2?

c) Tutki koordinaatistomallin avulla, millä todennäköisyydellä saadaan ensimmäisellä heitolla silmäluku 3 ja toisella heitolla silmäluku 2.

d) Mikä yhteys a- ja b-kohtien todennäköisyyksillä näyttäisi olevan c-kohdan todennäköisyyteen?



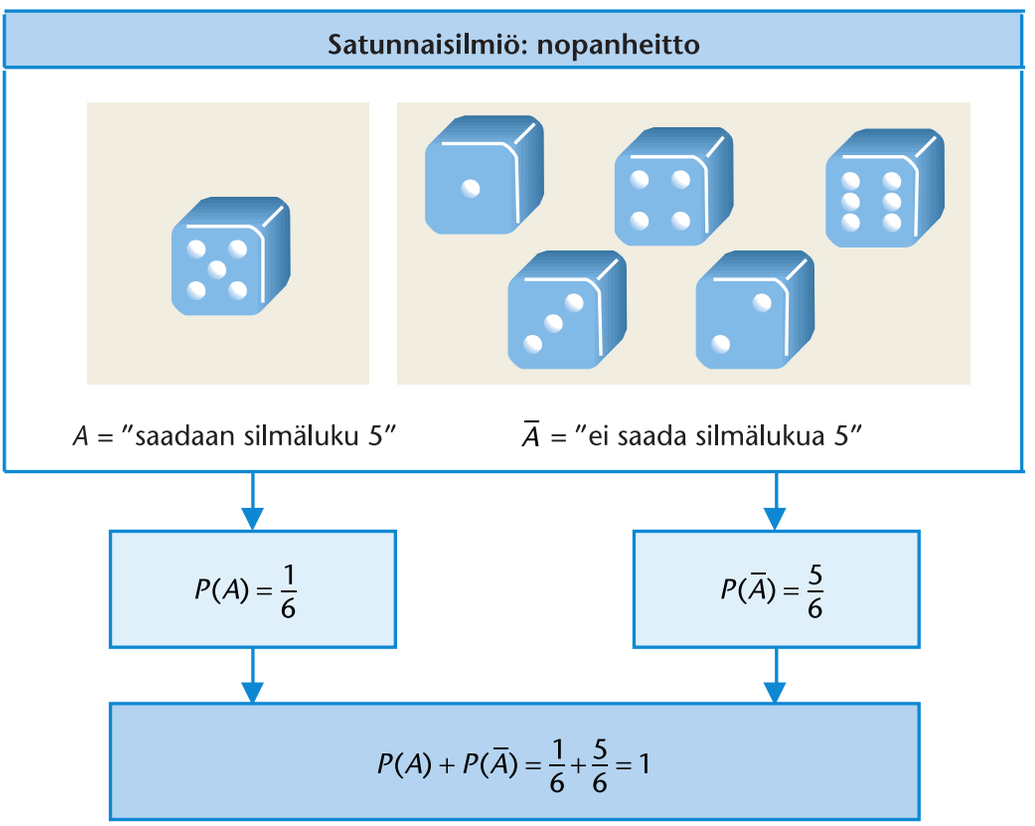
Komplementti

Jos satunnaisilmiö sisältää tapahtuman A, se sisältää myös A:n vastatapahtuman ”ei A”. Tätä vastatapahtumaa kutsutaan A:n **komplementtitapahtumaksi** ja sitä merkitään symbolilla $\overline{A}$.

Esimerkiksi satunnaisilmiö nopanheitto sisältää tapahtumat:

- A = ”saadaan silmäluku 5”
- $\overline{A}$ = ”ei saada silmälukua 5”

Jokainen perusjoukon tapaus kuuluu joko tapahtumaan A tai tapahtumaan $\overline{A}$, joten tapahtuman ja vastatapahtuman todennäköisyyksien summa on 1.



Tapahtuman A todennäköisyys voidaan laskea sen komplementtitapahtuman $\overline{A}$ todennäköisyyden avulla. Todennäköisyys kannattaa laskea komplementtitapahtuman avulla, jos vastatapahtuman todennäköisyyden laskeminen on oleellisesti helpompi ja lyhyempi tapa ratkaista tehtävä.

Todennäköisyys komplementin avulla

$$P(A) = 1 - P(\overline{A})$$

Kertolasku-sääntö

Satunnaisilmiön tapahtumat A ja B ovat **riippumattomia**, jos tapahtumien todennäköisyyksiin ei vaikuta se, tapahtuuko myös toinen vai ei. Voidaan ajatella, ettei tapahtumien välillä ole vuorovaikutusta.

Tarkastellaan tilannetta, jossa maljakossa on kaksi punaista, kaksi valkoista ja kaksi keltaista ruusua. Maljakosta nostetaan ensin satunnaisesti yksi ruusu ja palautetaan ruusu noston jälkeen takaisin maljakkoon. Toistetaan nosto. Tapahtumat ovat riippumattomia, koska ruusu palautetaan noston jälkeen takaisin maljakkoon.

Kootaan taulukkoon kaikki mahdolliset ruusuparit, joka kahdella nostolla voidaan saada:

Todennäköisyys, että molemmilla nostoilla saadaan punainen ruusu, on $\frac{1}{9}$.

Edellä tarkastellun tapahtuman ”saadaan molemmilla nostoilla punainen ruusu”, voidaan ajatella muodostuvan kahdesta riippumattomasta tapahtumasta.

NOSTO 1
A = ”saadaan punainen ruusu”
$P(A) = \frac{2}{6} = \frac{1}{3}$

NOSTO 2
B = ”saadaan punainen ruusu”
$P(B) = \frac{2}{6} = \frac{1}{3}$

Kahden riippumattoman tapahtuman todennäköisyyksien tulo

$$P(A) \cdot P(B) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9} = P(A \text{ ja } B)$$

Jos tilanne muodostuu kahdesta tai useammasta toisistaan riippumattomasta tapahtumasta, todennäköisyys saadaan näiden tapahtumien todennäköisyyksien tulona. Tämä tunnetaan kertolaskusääntönä.

Riippumattomien tapahtumien *A* ja *B* todennäköisyys

$P(A \text{ ja } B) = P(A) \cdot P(B)$

ESIMERKKI 1

Lentoyhtiön järjestämän kilpailun ensimmäinen palkinto on yhdistelmämatka kahteen kaupunkiin, jotka arvotaan. Millä todennäköisyydellä voittaja matkustaa vain Pohjoismaissa?

Kaupunki 1	Kaupunki 2
Tukholma	Reykjavik
Oslo	Pariisi
Lontoo	Ateena
Kööpenhamina	Rooma
Hampuri	
Moskova	

RATKAISU

Lasketaan ensin kummankin arvonnan kohdalla todennäköisyys sille, että matkakohde on Pohjoismaissa.

- $P(\text{ensimmäinen kaupunki on Pohjoismaissa}) = \frac{3}{6} = \frac{1}{2}$
- $P(\text{toinen kaupunki on Pohjoismaissa}) = \frac{1}{4}$

Kaupunkien arvonnat ovat toisistaan riippumattomia, joten kertolaskusäännön mukaan saadaan:

$$P(\text{ensimmäinen } \textbf{ja} \text{ toinen kaupunki ovat Pohjoismaissa})$$
$$= \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{8}$$

VASTAUS

$\frac{1}{8} = 0{,}125$

ESIMERKKI 2 Englannin kuullunymmärtämiskokeessa on monivalintatehtäviä, joissa vain yksi vaihtoehto neljästä on oikein. Opiskelija arvaa kuuden kysymyksen kokeessa satunnaisesti vastauksen kaikkiin kohtiin. Millä todennäköisyydellä

a) opiskelija vastaa kaikkiin kysymyksiin oikein

b) opiskelijan vastauksissa on ainakin yksi oikea vastaus?

RATKAISU

a) Todennäköisyys, että opiskelija vastaa yhteen kysymykseen oikein on $\frac{1}{4}$.

$$\begin{aligned} &P(\text{kaikki kuusi ovat oikein}) \\ &= P(\text{1. oikein } \textbf{ja} \text{ 2. oikein } \textbf{ja} \ \ldots \ \textbf{ja} \text{ 6. oikein}) \\ &= \underbrace{\frac{1}{4} \cdot \frac{1}{4} \cdot \ldots \cdot \frac{1}{4}}_{6 \text{ kpl}} \\ &= \left(\frac{1}{4}\right)^6 \\ &= 0{,}000244\ldots \\ &\approx 0{,}00024 \end{aligned}$$

b) Todennäköisyyden laskeminen suoraan on hankalaa, koska annetun ehdon, ”ainakin yksi oikea vastaus”, toteuttaa lukematon määrä rivejä. Opiskelijalla voisi olla ensimmäinen tehtävä oikein ja loput väärin, ensimmäinen ja kolmas tehtävä oikein ja loput väärin jne.

Tapahtuman komplementtitapahtuma on ”ei yhtään oikein” eli kaikki kuusi vastausta ovat väärin. Lasketaan kysytty todennäköisyys komplementtitapahtuman avulla.

Todennäköisyys arvata yksi kohta väärin on $\frac{3}{4}$.

$$P(\text{kaikki kuusi ovat väärin}) = \left(\frac{3}{4}\right)^6$$

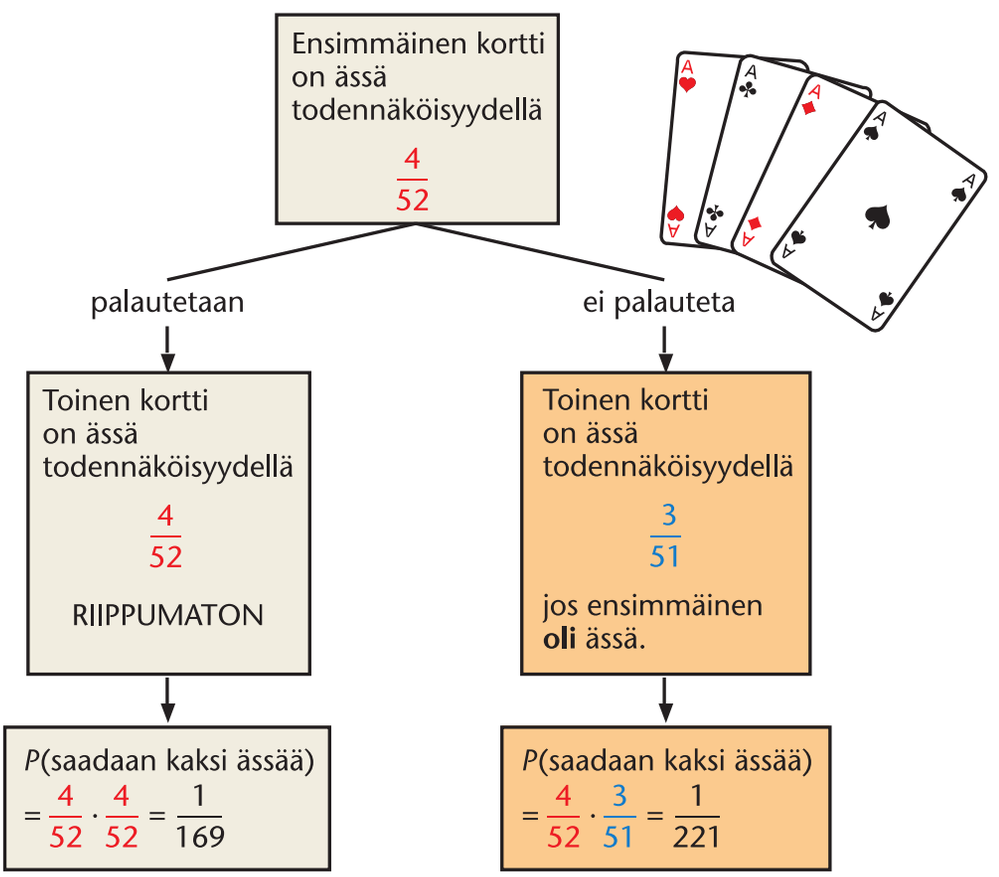
$$\begin{aligned} P(\text{ainakin yksi on oikein}) &= 1 - P(\text{kaikki kuusi ovat väärin}) \\ &= 1 - \left(\frac{3}{4}\right)^6 \\ &= 0{,}8220\ldots \\ &\approx 0{,}82 \end{aligned}$$

VASTAUS a) $0{,}0024 = 0{,}24\ \%$ b) $0{,}82 = 82\ \%$

Tapahtumat ovat riippumattomia esimerkiksi silloin, kun nostetaan kaksi korttia korttipakasta, mutta ensimmäinen kortti palautetaan noston jälkeen takaisin pakkaan, ja pakka sekoitetaan uudelleen. Tällöin esimerkiksi todennäköisyys saada ässä toisella nostolla on aina sama riippumatta siitä, mikä kortti ensimmäisellä kerralla nostettiin.

Jos korttia ei palauteta pakkaan, todennäköisyys saada toisella nostolla ässä riippuu ensimmäisen noston tuloksesta.

Kertolaskusääntöä voi soveltaa myös tilanteeseen, jossa tapahtumat riippuvat toisistaan. Tällöin peräkkäisten tapahtumien todennäköisyydet eivät välttämättä ole yhtä suuret. Yksittäisten tapahtumien todennäköisyyksien määrittämisessä on aina huomioitava, mitä aiemmin on jo tapahtunut.

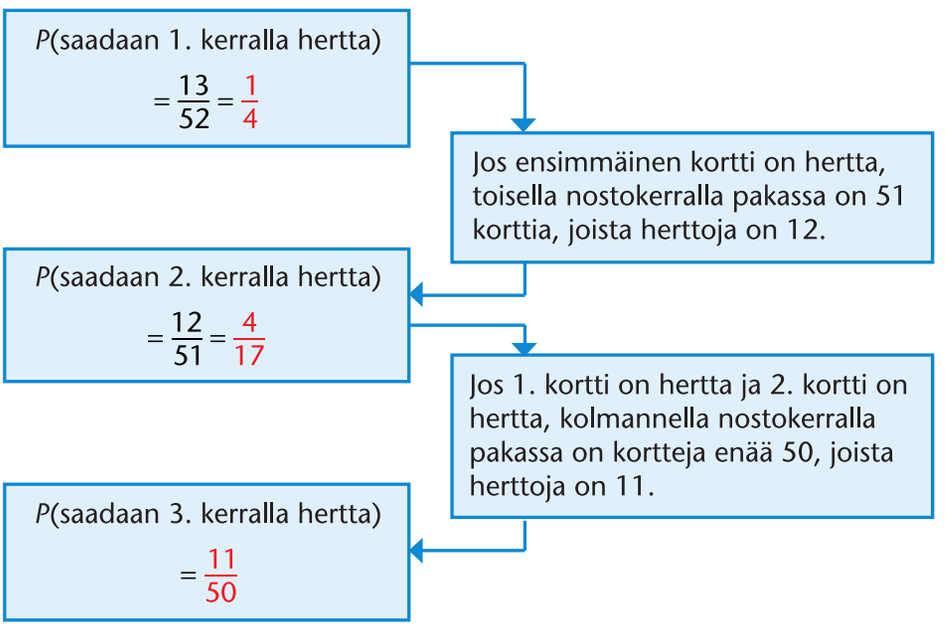


ESIMERKKI 3 Korttipakasta vedetään kolme korttia, joita ei palauteta pakkaan. Millä todennäköisyydellä

a) saadaan kolme herttaa b) ainakin yksi hertta?

RATKAISU

a) Tapahtuma "saadaan kolme herttaa" muodostuu kolmesta peräkkäisestä tapahtumasta. Pakassa on alussa 52 korttia, joista 13 on herttoja.



Tapahtuman ”1. kortti on hertta **ja** 2. kortti on hertta **ja** 3. kortti on hertta” todennäköisyys on

$$P(\text{saadaan kolme herttaa}) = \frac{1}{4} \cdot \frac{4}{17} \cdot \frac{11}{50} = \frac{11}{850} = 0{,}0129... \approx 0{,}013$$

b) Tapahtuman A = ”saadaan ainakin yksi hertta” todennäköisyys on helpompi laskea komplementtitapahtuman avulla:

$\overline{A}$ = ”ei saada yhtään herttaa”
= ”saadaan jokaisella kerralla joku muu kuin hertta”

Muita kortteja kuin herttoja on pakassa
52 – 13 = 39

- $P(\text{1. kortti on joku muu kuin hertta}) = \frac{39}{52} = \frac{3}{4}$
- $P(\text{2. kortti on joku muu kuin hertta}) = \frac{38}{51}$
- $P(\text{3. kortti on joku muu kuin hertta}) = \frac{37}{50}$

$$P(A) = 1 - P(\overline{A}) = 1 - \frac{3}{4} \cdot \frac{38}{51} \cdot \frac{37}{50} = 1 - \frac{703}{1700} = 0{,}5864... \approx 0{,}59$$

VASTAUS a) 0,013 b) 0,59

ESIMERKKI 4

Kaapissa on kirkkaita ja sinisiä Kartio-laseja yhteensä yhdeksän kappaletta. Juuso ottaa kaapista kaksi lasia. Molemmat lasit ovat sinisiä todennäköisyydellä $\frac{5}{12}$. Kuinka monta sinistä lasia kaapissa on?

RATKAISU

Merkitään sinisten lasien lukumäärää kirjaimella x.

$$P(\text{ensimmäinen lasi on sininen}) = \frac{x}{9}$$

Tämän jälkeen kaapissa on kahdeksan lasia, joista $x-1$ kappaletta on sinisiä.

$$P(\text{toinen lasi on sininen}) = \frac{x-1}{8}$$

Kertolaskusäännön mukaan todennäköisyys

$$P(\text{molemmat lasit ovat sinisiä}) = \frac{x}{9} \cdot \frac{x-1}{8} = \frac{x^2 - x}{72}$$

Koska tapahtuman ”molemmat lasit ovat sinisiä” todennäköisyys on $\frac{5}{12}$, saadaan yhtälö

$$\frac{x^2 - x}{72} = \frac{5}{12}$$
$$12x^2 - 12x = 360$$
$$12x^2 - 12x - 360 = 0$$

Ratkaisukaavalla saadaan

$$x = \frac{-(-12) \pm \sqrt{(-12)^2 - 4 \cdot 12 \cdot (-360)}}{2 \cdot 12}$$
$$x = \frac{12 \pm 132}{24}$$
$$x = 6 \quad \text{tai} \quad x = -5$$

$ax^2 + bx + c = 0$

$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$

Koska lukumäärä on positiivinen, ratkaisuista kelpaa vain $x = 6$.

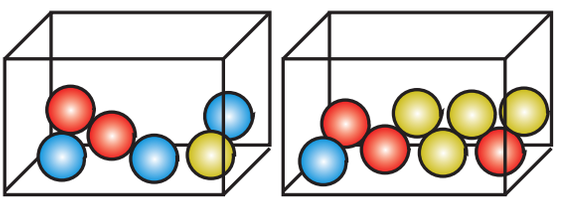
VASTAUS

Sinisiä laseja on kuusi kappaletta.

Tehtäviä

129. Säätiedotuksen mukaan lauantaina ukkostaa todennäköisyydellä 0,4 ja sunnuntaina todennäköisyydellä 0,3. Laske todennäköisyys, että viikonloppuna ukkostaa molempina päivinä.

E1 130. Taikurilla on kaksi laatikkoa, joissa on eriväristä palloja. Millä todennäköisyydellä taikuri nostaa kummastakin laatikosta punaisen pallon?



laatikko 1 laatikko 2

131. Ilari ja Matti harrastavat darts-tikkapeliä. Ilari osuu häränsilmään todennäköisyydellä 0,25 ja Matti todennäköisyydellä 0,18.
   a) Molemmat heittävät yhden tikan. Millä todennäköisyydellä molemmat heittävät häränsilmään?
   b) Matti heittää kolme kertaa. Millä todennäköisyydellä hän ei osu kertaakaan häränsilmään?

132. Noppaa heitetään neljä kertaa. Millä todennäköisyydellä saadaan
   a) kaikilla kerroilla kuusi
   b) kaikilla kerroilla vähintään viisi?

133. Verkkopankkiin kirjaudutaan niin, että ensin annetaan kuuden numeron pituinen käyttäjätunnus ja neljän numeron pituinen salasana, minkä jälkeen annetaan vielä neljän numeron pituinen kertakäyttötunnus. Jokaisella pankin asiakkaalla on eri käyttäjätunnus, mutta usealla asiakkaalla voi olla sama salasana ja kertakäyttötunnus.
   a) Jos verkkopankilla on 600 000 asiakasta, niin mikä on todennäköisyys sille, että yhdellä arvauksella löytää jonkun asiakkaan käyttäjätunnuksen?
   b) Mikä on todennäköisyys sille, että yhdellä yrityksellä pääsee kirjautumaan verkkopankkiin? [S2010, 8]

E2 134. Pankkikortin tunnusluvussa on neljä numeroa, jotka voivat olla mitä tahansa numeroita 0–9. Vilma on unohtanut oman tunnuslukunsa. Millä todennäköisyydellä hän umpimähkään kokeillessaan näppäilee
a) oikean luvun
b) ainakin yhden oikean numeron?

135. Noin viidesosa Suomen asukkaista asuu pääkaupunkiseudulla. Tietokone arpoo kuusi suomalaista. Millä todennäköisyydellä tässä joukossa on ainakin yksi pääkaupunkiseudulla asuva henkilö?

136. Kadun varrella on kymmenen maksutonta parkkipaikkaa. Kukin paikka on vapaana keskimäärin kuusi minuuttia tunnissa. Kuinka suuri on todennäköisyys, että satunnaisesti paikalle saapuva autoilija
a) ei löydä vapaata parkkipaikkaa
b) löytää vapaan parkkipaikan?

137. Kukkakaupassa myydään tulppaanin sipuleita kuuden kappaleen pakkauksissa. Itämistodennäköisyys on 0,8. Millä todennäköisyydellä
a) kaikki pakkauksen sipulit itävät
b) ainakin yksi sipuleista itää?

138. Tiedonsiirtojärjestelmässä havaittiin yksittäisen bitin saapuvan virheellisenä vastaanottajalle todennäköisyydellä 0,00015. Yksittäisten bittien siirtojen oletetaan olevan toisistaan riippumattomia.
a) Millä todennäköisyydellä vastaanottajalle saapuvassa 16 bitin jonossa on ainakin yksi virheellinen bitti?
b) Jos lähetetään 32 kappaletta 16 bitin jonoja, niin millä todennäköisyydellä vastaanottajalle saapuu ainakin yksi virheellinen jono? [K2010, 8]

139. Erään väestötilaston mukaan 18-vuotiaiden tyttöjen pituuden keskiarvo on 165 cm ja keskihajonta 6 cm. Samanikäisten tyttöjen pituus noudattaa likimain normaalijakaumaa. Mikä on todennäköisyys sille, että umpimähkään valittu 18-vuotias tyttö on korkeintaan 175 cm pitkä? Millä todennäköisyydellä kolmen tytön ryhmässä kaikki ovat alle 175 cm pitkiä? Entä millä todennäköisyydellä kolmesta tytöstä ainakin yksi on yli 175 cm pitkä? [S2009, 9]

E3 140. Korissa on 8 vihreää omenaa ja 17 punaista omenaa. Korista otetaan neljä omenaa piirakan tekoa varten. Millä todennäköisyydellä
a) kaikki omenat ovat punaisia
b) ainakin yksi omenoista on vihreä
c) ainakin yksi omenoista on punainen?

141. Äänilevykokoelmassa on 12 levyä, joista viisi on sinfonia-, neljä jazz- ja kolme oopperamusiikkia. Liisa valitsee umpimähkään kaksi levyä. Millä todennäköisyydellä molemmat ovat jazzmusiikkia?

142. Ventti on korttipeli, jossa pyritään saamaan kolmella kortilla yhteispistemääräksi 21. Millä todennäköisyydellä saadaan ventti kolmella seiskalla?

143. Korttipakasta nostetaan viisi korttia, joita ei palauteta pakkaan. Millä todennäköisyydellä nostetuista korteista ainakin yksi on
a) kuvakortti b) risti?

144. Arpajaisissa on myynnissä 50 arpaa, joista 30 on voittoarpoja. Rasmus ostaa arpajaisista neljä arpaa. Millä todennäköisyydellä
a) kaikki arvat ovat voittoarpoja
b) mikään arvoista ei ole voittoarpa
c) ainakin yksi arvoista on voittoarpa
d) ainakin yksi arvoista on sellainen, jolla ei voita mitään?

E4 145. Sari ostaa torilta korillisen kirsikoita. Hän poimii 50 kirsikan joukosta satunnaisesti kaksi kirsikkaa. Todennäköisyys, että molemmat ovat pilaantuneita, on $\frac{3}{175}$. Kuinka monta pilaantunutta kirsikkaa Sari on ostanut?

146. Lauri ostaa kaksi arpaa. Oletetaan, että jokaisella yksittäisellä arvalla on sama todennäköisyys voittaa. Laske tämä todennäköisyys, kun todennäköisyys sille, että Laurin molemmat arvat voittavat, on 0,34.

147. Korttipakasta on hävinnyt kortteja. Kaikki puuttuvat kortit ovat patoja. Pakasta vedetään kaksi korttia, joita ei palauteta pakkaan. Todennäköisyys saada kaksi pataa on $\frac{3}{94}$. Laske tämän avulla, kuinka monta pataa pakasta on kadonnut.

148. Vasenkätisiä on erään tiedon mukaan 10 % väestöstä. Kuinka monta henkilöä tulee satunnaisesti kootussa ryhmässä olla, jotta siinä olisi ainakin yksi vasenkätinen todennäköisyydellä 0,8? [K2009, 11]

Sekalaisia tehtäviä

149. Lotan koulumatkan varrella on kolmet liikennevalot. Ensimmäiset näyttävät vihreää keskimäärin 40 min tunnista, toiset 35 min tunnista ja kolmannet 28 min tunnista. Millä todennäköisyydellä Lotta ei joudu pysähtymään liikennevaloihin matkallaan kouluun?

150. Raha-automaattiyhdistyksen pokeripeleissä saadun voittosumman voi ”tuplata” lyömällä vetoa siitä, onko jokerittomasta pakasta umpimähkään nouseva kortti pieni (1–6) vai suuri (8–13). Pelaaja häviää, jos nouseva kortti on 7. Jos pelaaja arvaa oikein, hän voi jatkaa arvaamista tai lunastaa voiton.
   a) Millä todennäköisyydellä pelaaja voittaa, kun hän tuplaa viisi kertaa ja lyö vetoa aina suuren kortin puolesta?
   b) Millä todennäköisyydellä pakasta nousee viisi kertaa 7?

151. Metsänhoitaja tietää kokemuksesta, että keskimäärin yksi talvi viidestä aiheuttaa epäedullisten sääolosuhteiden takia vahinkoja taimikoille. Hän tietää myös, että hirvet tuhoavat taimikkoa keskimäärin kahdesti kuuden talven aikana. Millä todennäköisyydellä taimikko joutuu ensi talvena ainakin jommankumman tuhon kohteeksi?

152. Puutarhuri istuttaa siemeniä, joiden itävyys on 60 %.
   a) Mikä on todennäköisyys, että kolmesta istutetusta siemenestä mikään ei idä? Mikä on todennäköisyys, että ainakin yksi siemen itää?
   b) Siemeniä istutetaan viiteen ruukkuun kuhunkin kolme. Mikä on todennäköisyys, että jokaisessa ruukussa ainakin yksi siemen itää? [S2008, 11]

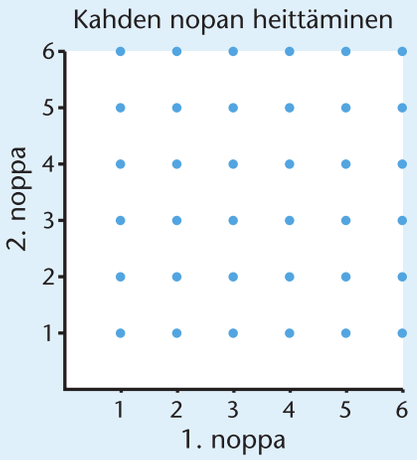
153. Korttipakasta nostetaan neljä korttia. Millä todennäköisyydellä saadaan neljä ässää, kun
   a) kortti palautetaan noston jälkeen pakkaan
   b) nostettua korttia ei palauteta pakkaan?

154. Eräässä lukion opetusryhmässä on 15 tyttöä. Laske ryhmän poikien lukumäärä, kun umpimähkään valitut kaksi oppilasta ovat tyttöjä todennäköisyydellä $\frac{1}{6}$.

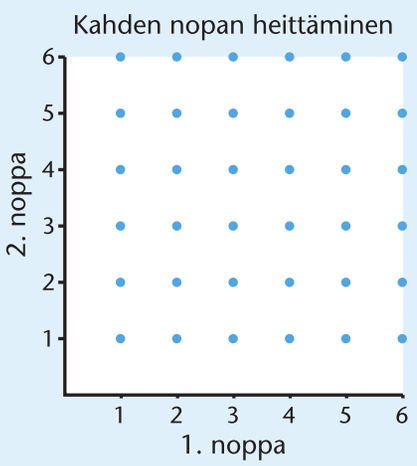
2.3 Yhteenlaskusääntö

POHDINTA 1 Heitetään kahta noppaa. Tutki koordinaatistomallin avulla, millä todennäköisyydellä

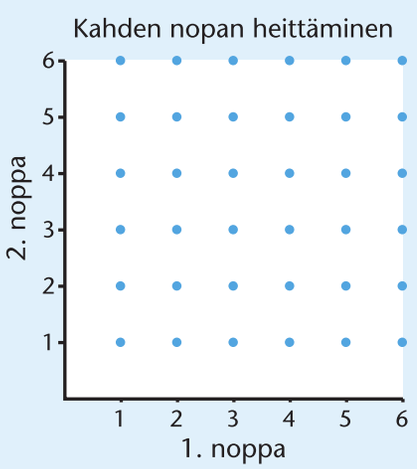
a) ensimmäisellä heitolla saadaan silmäluku 5, mutta toisella jokin muu



b) toisella heitolla saadaan silmäluku 5 ja ensimmäisellä jokin muu



c) vain toisella heitoista saadaan silmäluku 5.



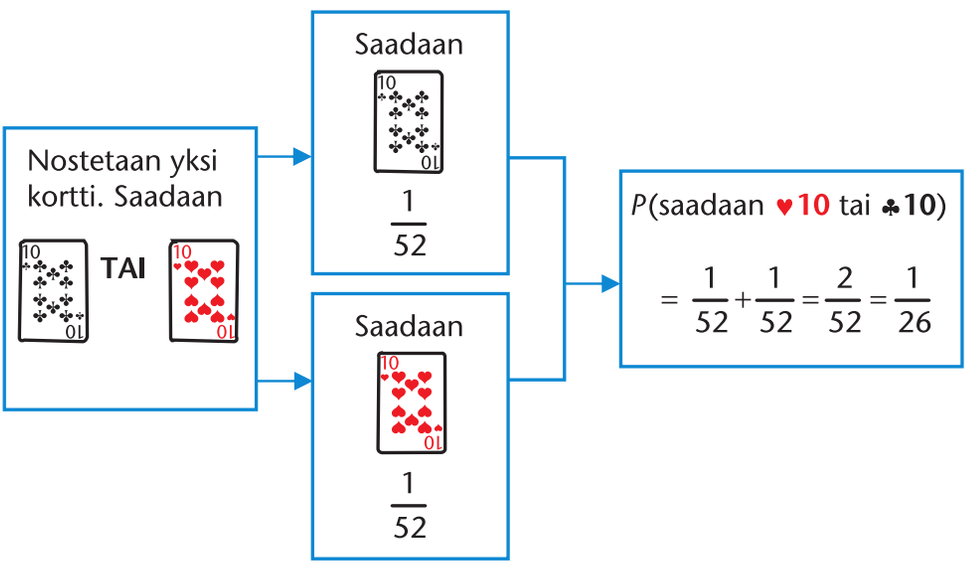
Laske a- ja b-kohtien todennäköisyydet yhteen.
Vertaa tätä c-kohdan todennäköisyyteen. Mitä huomaat?

Kahden peräkkäisen tapahtuman todennäköisyys saadaan kertolaskusäännön avulla. Tapahtumat voivat olla myös **vaihtoehtoisia**. Tällöin todennäköisyyden laskemiseen ei riitä kertolaskusääntö.

Tarkastellaan esimerkkinä kortin vetämistä tavallisesta 52 kortin pakasta. Vedetään pakasta yksi kortti. Tapahtuman ”saadaan nostettaessa herttakymppi ♥**10** **tai** ristikymppi ♣**10**” suotuisia tapauksia on vain kaksi. Tällöin

$$P(\text{saadaan } \heartsuit 10 \text{ tai } \clubsuit 10) = \frac{2}{52} = \frac{1}{26}$$

Tämän tapahtuman voidaan ajatella muodostuvan kahdesta vaihtoehtoisesta tapahtumasta, joille molemmille voidaan erikseen laskea todennäköisyys:



Tapahtuman ”saadaan nostettaessa pata ♠ **tai** kuningatar Q” suotuisia tapauksia on 16 (**12 pataa** + ♠Q, ♥Q, ♣Q ja ♦Q). Tällöin

$$P(\text{saadaan } \spadesuit \text{ tai } \mathbf{Q}) = \frac{16}{52} = \frac{4}{13}$$

Tämäkin tapahtuma koostuu kahdesta vaihtoehtoisesta tapahtumasta, ”saadaan pata” tai ”saadaan kuningatar”. Patakuningatar sisältyy näihin molempiin tapahtumiin. Vähentämällä patakuningattaren todennäköisyys päädytään samaan lopputulokseen kuin edellä:

$$P(\text{saadaan } \spadesuit \text{ tai } \mathbf{Q}) = P(\spadesuit) + P(\mathbf{Q}) - P(\spadesuit\mathbf{Q})$$
$$= \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{4}{13}$$

Patakuningatar huomioidaan vain kerran.

Satunnaisilmiöön liittyvät tapahtumat voivat olla **erillisiä** eli toisensa poissulkevia tai **ei-erillisiä**, jolloin toinen tapahtuma sisältää toisen kokonaan tai osittain.

Satunnaisilmiö	Tapahtumat	
kortin vetäminen	• saadaan ristikymppi • saadaan herttakymppi	erilliset tapahtumat
kortin vetäminen	• saadaan rouva • saadaan pata	toinen sisältää osittain toisen tapahtuman
kortin vetäminen	• saadaan patarouva • saadaan pata	toinen sisältää kokonaan toisen tapahtuman

Jos tapahtumat A tai B ovat erillisiä, tapahtuman ”A tai B” todennäköisyys saadaan laskemalla yhteen tapahtumien todennäköisyydet. Tämä tunnetaan **yhteenlaskusääntönä**.

Erillisten tapahtumien yhteenlaskusääntö

$$P(A \text{ tai } B) = P(A) + P(B)$$

Matematiikassa sana TAI tapahtumien A ja B välillä tarkoittaa, että tapahtuu jompikumpi tai molemmat eli ”A tai B tai molemmat”.

Jos tapahtumat A ja B eivät ole erillisiä, tapahtuman ”A tai B” todennäköisyys saadaan laskemalla tapahtumien A ja B todennäköisyydet yhteen ja vähentämällä tuloksesta yhteisen osan, (tapahtuman ”A ja B”) todennäköisyys.

Yleinen yhteenlaskusääntö

$$P(A \text{ tai } B) = P(A) + P(B) - P(A \text{ ja } B)$$

ESIMERKKI 1

Yrityksen henkilöstöstä 40 % käyttää työssään tablettia ja 70 % tavallista tietokonetta. Molempia laitteita käyttää 25 % työntekijöistä. Millä todennäköisyydellä satunnaisesti valittu työntekijä käyttää työssään tablettia tai tietokonetta?

RATKAISU

Tapa 1

Havainnollistetaan tehtävää Venn-diagrammin avulla.

- Harmaa alue kuvaa henkilöitä, jotka käyttävät molempia laitteita:

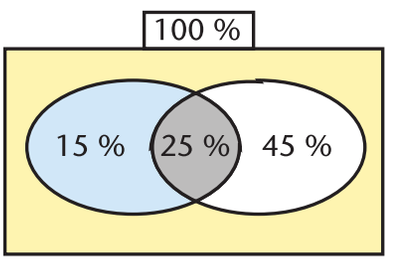
  25 %

- Sininen alue kuvaa henkilöitä, jotka käyttävät vain tablettia:

  $40\ \% - 25\ \% = \mathbf{15\ \%}$

- Valkoinen alue kuvaa henkilöitä, jotka käyttävät vain tavallista tietokonetta:

  $70\ \% - 25\ \% = \mathbf{45\ \%}$



Todennäköisyys sille, että henkilö käyttää tablettia tai tavallista konetta, saadaan **laskemalla kolmen alueen todennäköisyydet yhteen**:

$$15\ \% + 25\ \% + 45\ \% = 85\ \% = 0{,}85$$

Tapa 2

Tapahtumat eivät ole erillisiä, koska yrityksessä on henkilöitä, jotka käyttävät sekä tablettia että tietokonetta. Käytetään laskemiseen yleistä yhteenlaskusääntöä:

$$\begin{aligned} &P(\text{käyttää tablettia tai tavallista tietokonetta}) \\ &= P(\text{tablettia}) + P(\text{tavallista tietokonetta}) - P(\text{molempia}) \\ &= 0{,}4 + 0{,}7 - 0{,}25 \\ &= 0{,}85 \end{aligned}$$

VASTAUS 0,85

ESIMERKKI 2

Valtteri osuu koripallon vapaaheitossa koriin todennäköisyydellä 0,6. Hän heittää kolme kertaa. Millä todennäköisyydellä Valtteri osuu koriin

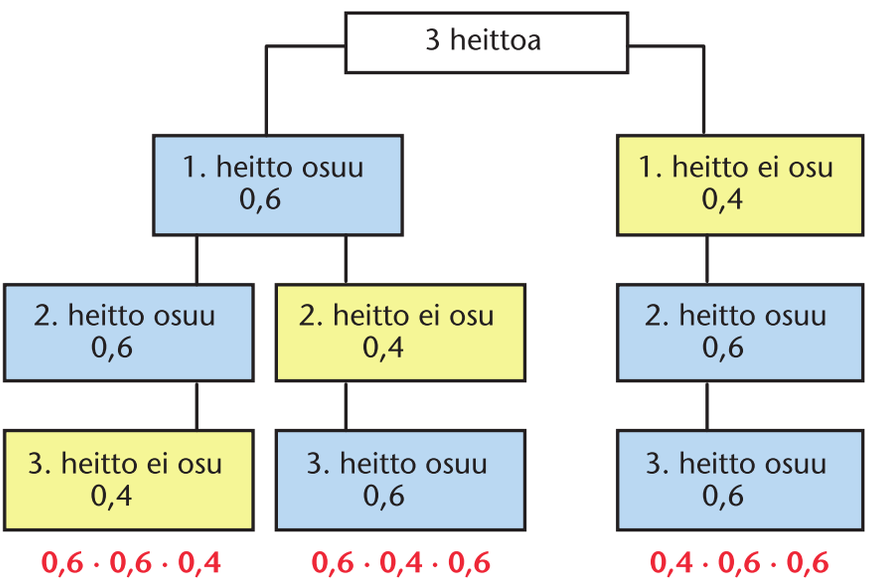
a) kahdella ensimmäisellä heitolla, mutta ei kolmannella
b) täsmälleen kahdella heitolla kolmesta?

RATKAISU

a) Koska Valtteri heittää korin todennäköisyydellä 0,6, hän heittää ohi todennäköisyydellä $1 - 0{,}6 = 0{,}4$. Heittosarjan ”1. osuu, 2. osuu ja 3. ei osu” todennäköisyys saadaan kertomalla peräkkäisten heittojen todennäköisyydet.

$$P(\text{osuu koriin 1. ja 2. heitolla, mutta ei 3. heitolla}) = 0{,}6 \cdot 0{,}6 \cdot 0{,}4 = 0{,}144$$

b) Suotuisassa heittosarjassa on kaksi heittoa koriin. Ne voivat olla mitkä tahansa kaksi heittoa kolmesta. Havainnollistetaan tilannetta puukaavion avulla. Jokainen puun haara kuvaa kolmen peräkkäisen heiton sarjaa. Yhden haaran todennäköisyys lasketaan siis kertolaskusäännön avulla.



Koska kunkin haaran todennäköisyys on tässä tilanteessa sama, yhden haaran todennäköisyys voidaan kertoa kolmella.

Kokonaistodennäköisyys on näiden kolmen eri heittosarjan todennäköisyyksien summa.

$$P(\text{osuu koriin kaksi kertaa}) = 3 \cdot 0{,}6 \cdot 0{,}6 \cdot 0{,}4 = 0{,}432$$

VASTAUS

a) 0,144 b) 0,432

ESIMERKKI 3

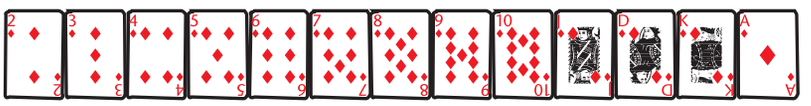
Tavallisesta 52 kortin pakasta vedetään kolme korttia palauttamatta niitä pakkaan. Millä todennäköisyydellä saadaan ainakin kaksi ruutua?

RATKAISU

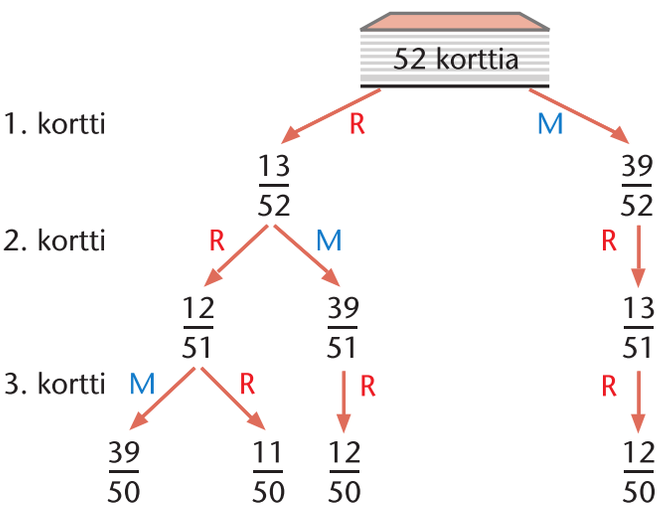
Tapahtuma ”saadaan ainakin kaksi ruutua” voi tarkoittaa seuraavia tilanteita:

- saadaan täsmälleen kaksi ruutua (**R**) ja yksi muu maa (**M**)
- saadaan kaikilla kolmella kortilla ruutu.

Pakassa on 13 ruutua. Muiden maiden kortteja on yhteensä 39.



Havainnollistetaan tilannetta puukaavion avulla. Korttien määrä pakassa vähenee yhdellä jokaisen noston jälkeen.



Lasketaan puun haarojen todennäköisyydet kertolaskusäännön avulla:

$$\begin{aligned}
&\text{RRM} && \frac{13}{52}\cdot\frac{12}{51}\cdot\frac{39}{50}=\frac{39}{850}\\
&\text{RMR} && \frac{13}{52}\cdot\frac{39}{51}\cdot\frac{12}{50}=\frac{39}{850}\\
&\text{MRR} && \frac{39}{52}\cdot\frac{13}{51}\cdot\frac{12}{50}=\frac{39}{850}\\
&\text{RRR} && \frac{13}{52}\cdot\frac{12}{51}\cdot\frac{11}{50}=\frac{11}{850}
\end{aligned}
\qquad 3\cdot\frac{39}{850}$$

Kokonaistodennäköisyys on haarojen todennäköisyyksien summa:

$$P(\text{saadaan ainakin kaksi ruutua}) = 3\cdot\frac{39}{850}+\frac{11}{850}=\frac{64}{425}\approx 0{,}15$$

VASTAUS

$\frac{64}{425}\approx 0{,}15$

ESIMERKKI 4

VR tutki junien myöhästymisiä lumen ja pakkasen takia Savon radan asemilla Kuopio – Pieksämäki – Mikkeli. Jos juna saapui asemalle aikataulun mukaisesti, se oli seuraavalla asemalla myöhässä todennäköisyydellä 0,10. Jos juna saapui asemalle myöhässä, se saapui myös seuraavalle asemalle myöhässä todennäköisyydellä 0,45. Millä todennäköisyydellä juna saapuu aikataulun mukaisesti Mikkeliin, kun se oli myöhässä Kuopiossa?

RATKAISU

Kuopion aseman jälkeen juna voi saapua Pieksämäelle joko aikataulun mukaisesti tai myöhässä. Tämän tapahtuman todennäköisyys vaikuttaa seuraavaan asemaan, Mikkeliin.

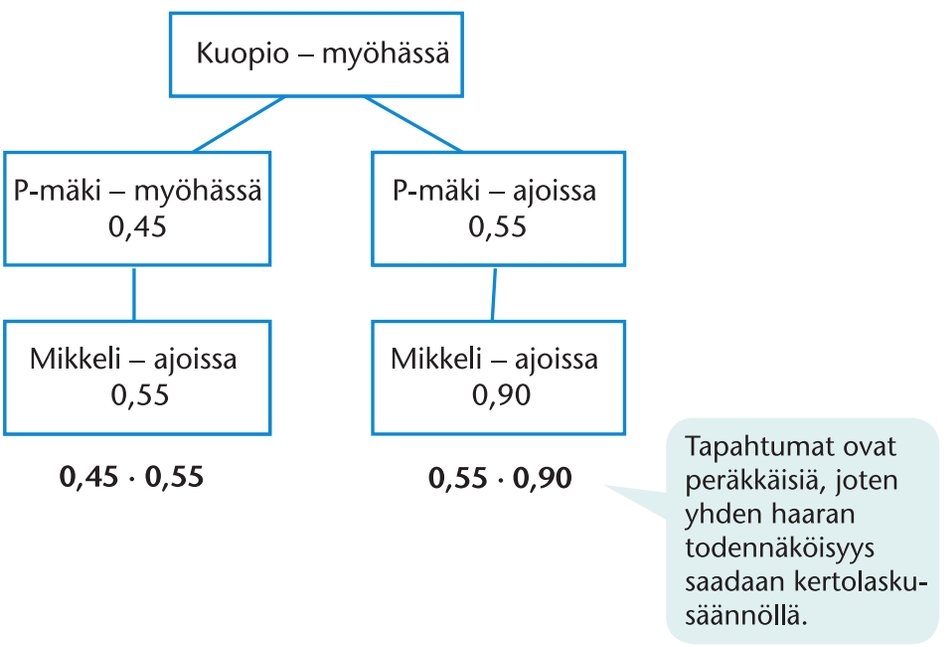
Jos juna saapuu asemalla aikataulun mukaisesti, seuraavalla asemalla se on

- myöhässä todennäköisyydellä 0,10
- ajoissa todennäköisyydellä $1 - 0{,}10 = 0{,}90$.

Jos juna saapuu asemalle myöhässä, seuraavalla asemalla se on

- myöhässä todennäköisyydellä 0,45
- ajoissa todennäköisyydellä $1 - 0{,}45 = 0{,}55$.

Piirretään puukaavio havainnollistamaan tilannetta.



Lasketaan todennäköisyys yhteenlaskusäännön avulla:

$$P(\text{Mikkeliin aikataulun mukaisesti}) = 0{,}45 \cdot 0{,}55 + 0{,}55 \cdot 0{,}90 = 0{,}7425 \approx 0{,}74$$

VASTAUS 0,74

Tehtäviä

155. Mitkä seuraavista tapahtumista ovat erillisiä tapahtumia eli tapahtumia, jotka eivät sisällä toisiaan?
   a) Silmäluvun 4 saaminen nopanheitossa ja kruunan saaminen kolikonheitossa.
   b) 17-vuotiaat ja 18-vuotiaat.
   c) 17-vuotiaat lukiolaiset ja 18-vuotiaat lukiolaiset.
   d) Abiturientit ja lukiolaiset.
   e) Vihreäsilmäiset ihmiset ja sinisilmäiset ihmiset.
   f) Ristin vetäminen korttipakasta ja ristiässän vetäminen korttipakasta.

156. Korttipakasta vedetään yksi kortti. Millä todennäköisyydellä
   a) kortti on hertta tai pata b) kortti on hertta tai ässä?

E1 157. Noin 33 % suomalaisista kuuluu veriryhmään O ja noin 13 %:lla on negatiivinen Rhesus-tekijä. Henkilöitä, joiden veriryhmä on O ja Rhesus-tekijä negatiivinen, on vain 5 %. Millä todennäköisyydellä satunnaisesti valittu suomalainen kuuluu veriryhmään O tai hänellä on negatiivinen Rhesus-tekijä?

158. Kennelin koirista 70 % on lyhytkarvaisia ja loput pitkäkarvaisia. Lyhytkarvaisista 25 % on mustia. Kennelissä ei ole yhtään mustaa, pitkäkarvaista koiraa. Lyhytkarvaisista puolet ja pitkäkarvaisista 60 % on kääpiökoiria. Millä todennäköisyydellä satunnaisesti valittu kennelin koira on
   a) musta tai pitkäkarvainen
   b) kääpiökoira
   c) pitkäkarvainen tai kääpiökoira?

E2 159. Ominaisuus A periytyy jälkeläiselle todennäköisyydellä 0,15 ja ominaisuus B todennäköisyydellä 0,65. Ominaisuudet periytyvät toisistaan riippumattomina. Millä todennäköisyydellä jälkeläinen perii
   a) ominaisuuden A, mutta ei ominaisuutta B
   b) perii vain toisen ominaisuuksista A tai B?

E3 160. 8,0 % suomalaisista miehistä ei erota punaista ja vihreää väriä. Millä todennäköisyydellä kolmen satunnaisesti valitun miehen joukossa on ainakin kaksi punavihersokeaa?

161. Heitetään kolmesti lanttia. Millä todennäköisyydellä saadaan yksi klaava?

162. Viisi poikaa ja kaksi tyttöä saapuvat odottamaan hissiä. Hissin saapuessa osoittautuu, että siihen mahtuu vain kaksi henkilöä. Millä todennäköisyydellä
a) molemmat tytöt pääsevät hissiin
b) vain toinen tytöistä pääsee hissiin?

163. Ostetaan lasinen maljakko. On todettu, että 12 %:ssa maljakoista on valmistusvika. Vian vuoksi puolet näistä maljakoista on saanut myös kuljetusvaurion. Muista maljakoista vain 10 % on saanut kuljetusvaurion. Laske todennäköisyys, että ostamassasi maljakossa
a) on ainakin toinen vioista
b) ei ole mitään vikaa.

164. Opiskelija kulkee kouluun kolmen jalankulkuvalon kautta. Vihreän valon todennäköisyys on ensimmäisissä valoissa 0,80, toisissa 0,70 ja viimeisissä 0,90. Millä todennäköisyydellä
a) opiskelijan ei tarvitse pysähtyä valoissa kertaakaan
b) opiskelija joutuu pysähtymään vain kerran?

165. Abiturientilla on kaksi herätyskelloa. Uudempi toimii oikein 98 %:n todennäköisyydellä ja vanhempi 85 %:n todennäköisyydellä. Matematiikan kokeen aattona abiturientti asettaa molemmat kellot soimaan seuraavana aamuna. Millä todennäköisyydellä kelloista soi oikeaan aikaan
a) molemmat
b) vain toinen
c) ei kumpikaan?
[S2006, 4]

166. Minttu ostaa irtokarkkeja. Paperipussissa on seitsemän liitulakua, kuusi merirosvorahaa ja kahdeksan suklaarusinaa. Minttu ottaa pussista umpimähkään kaksi karkkia. Laske todennäköisyys, että
a) molemmat ovat liitulakuja
b) ainakin toinen on merirosvoraha.

E4 167. Koulusta myöhästynyt oppilas myöhästyy myös seuraavana päivänä 30 % todennäköisyydellä. Jos oppilas on tullut kouluun ajoissa, hän myöhästyy seuraavana päivänä 10 % todennäköisyydellä. Kuinka suurella todennäköisyydellä oppilas tulee keskiviikkona ajoissa kouluun, jos hän myöhästyi koulusta saman viikon maanantaina?

168. Silmien erivärisyys eli heterokromia voi johtua perinnöllisistä syistä. Eräässä suvussa esiintyy tätä ominaisuutta. Jos vanhemmalla on heterokromia, myös lapsi saa ominaisuuden todennäköisyydellä 0,25. Jos vanhemmalla ei ole tätä ominaisuutta, lapsella on ominaisuus todennäköisyydellä 0,12. Millä todennäköisyydellä heterokromisen naisen lapsenlapsella on heterokromia?

Sekalaisia tehtäviä

169. Pakasta vedetään kaksi korttia, joita ei palauteta pakkaan. Millä todennäköisyydellä vain toinen korteista on hertta?

170. Englannin kuuntelukokeessa on monivalintatehtäviä, joista yksi neljästä vaihtoehdosta on oikea. Opiskelija joutuu arvaamaan vastaukseen kolmeen ensimmäiseen kohtaan. Millä todennäköisyydellä
a) ensimmäinen menee oikein, mutta kaksi seuraavaa väärin
b) kaksi ensimmäistä menevät oikein, mutta kolmas väärin
c) hän arvaa oikein täsmälleen kaksi kolmesta?

171. Laatikossa on kolme mustaa, viisi valkoista ja neljä punaista noppaa. Nostetaan umpimähkään yksi noppa ja heitetään se. Millä todennäköisyydellä nostetaan musta noppa ja heitetään sillä kuutonen tai nostetaan valkoinen noppa ja heitetään sillä ykkönen?

172. Pituushyppääjä astuu hyppynsä yli 15 %:n todennäköisyydellä, jos hänen edellinen hyppynsäkin oli astuttu yli. Jos hyppy onnistuu, seuraava hyppy astutaan yli enää 10 %:n todennäköisyydellä. Hyppääjä astuu ensimmäisen hyppynsä yli. Millä todennäköisyydellä
a) kaksi seuraavaakin hyppyä astutaan yli
b) kolmas hyppy on yliastuttu?

2.4 Tuloperiaate ja kombinaatio

POHDINTA 1

Tarkastellaan henkilöiden A, B, C ja D muodostamaa ryhmää.

a) Asetetaan henkilöt jonoon. Kuinka monta erilaista jonoa voidaan muodostaa?

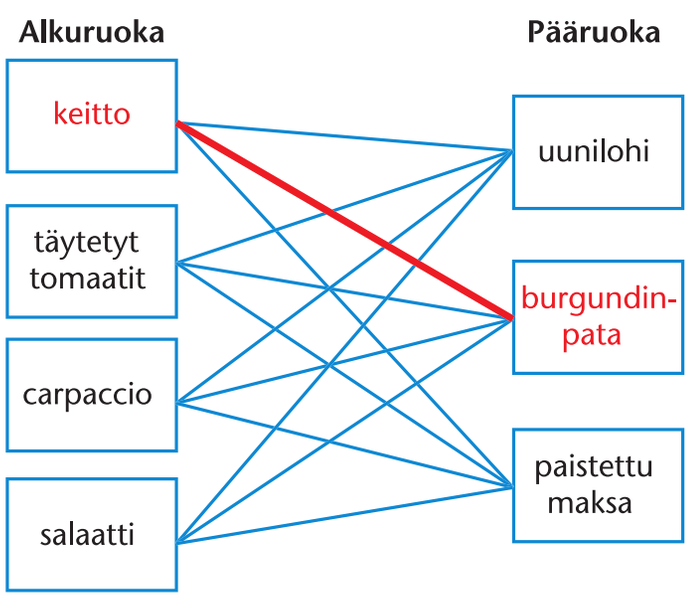
b) Ryhmästä muodostetaan pareja. Kuinka monta erilaista paria voidaan muodostaa?

c) Henkilöiden kesken arvotaan kaksi matkalahjakorttia, joiden arvot ovat 1000 € ja 500 €. Kuinka monella tavalla voittajat voidaan arpoa?

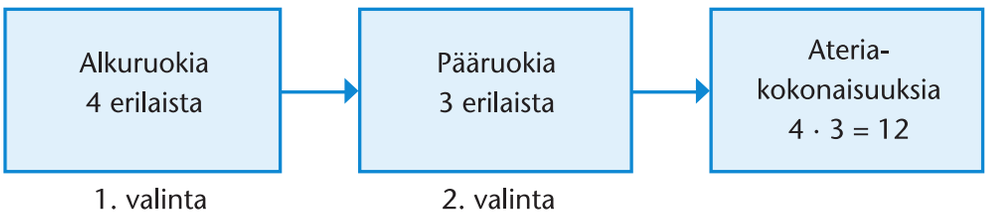
Tuloperiaate

Kokonaisuus voi muodostua useiden peräkkäisten valintojen kautta. Tarkastellaan esimerkkinä ravintolan lounaslistaa, josta asiakas voi valita neljän alkuruoan joukosta yhden ja kolmen pääruoan joukosta yhden.

Yksi ateriakokonaisuus muodostuu, kun asiakas valitsee alkuruoaksi esimerkiksi keiton ja pääruoaksi burgundinpataa. Uusi ateriakokonaisuus saadaan vaihtamalla burgundinpata vaikkapa uuniloheen. Näitä ateriakokonaisuuksia voi muodostaa listalta 12 erilaista:



Ratkaisun idea tunnetaan **tuloperiaatteena**. Jos kokonaisuus muodostuu usean peräkkäisen vaiheen kautta ja jokaisessa vaiheessa voidaan suorittaa tietty määrä valintoja, on erilaisten kokonaisuuksien määrä laskettavissa kertolaskun avulla:



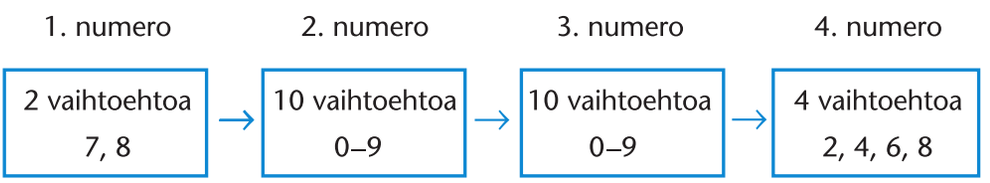
Vaiheita, joissa valintoja suoritetaan, voi olla useampia kuin kaksi. Jos kyseisen ravintolan listalla olisi vielä valittavana kaksi jälkiruokaa, erilaisia ateriakokonaisuuksia olisi yhteensä $\mathbf{4 \cdot 3 \cdot 2 = 24}$.

ESIMERKKI 1

Useissa kerrostaloissa on ovikoodi eli laite, josta valitaan neljä numeroa oven avaamiseksi. Pauliina on menossa ystävänsä luo, mutta hän on unohtanut oikean koodin. Hän muistaa, että koodin ensimmäinen numero on 7 tai 8 ja viimeinen nollasta eroava parillinen luku. Millä todennäköisyydellä ovi aukeaa?

RATKAISU

Ovikoodin valitseminen koostuu neljästä erillisestä vaiheesta.



Tuloperiaatteen mukaan erilaisia numerosarjoja on siis

$$2 \cdot 10 \cdot 10 \cdot 4 = 800$$

Niistä vain yksi on oikea eli

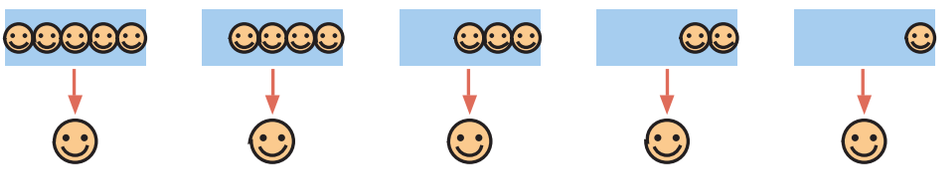
$$P(\text{Pauliina näppäilee oikean numerosarjan}) = \frac{1}{800} = 0{,}00125$$

VASTAUS

Pauliina saa oven auki todennäköisyydellä $\frac{1}{800} = 0{,}00125$.

Tuloperiaatetta voidaan soveltaa, kun muodostetaan erilaisia jonoja. Esimerkiksi viisi henkilöä voi asettua jonoon seuraavasti:

- Ensimmäiselle paikalle voi mennä kuka tahansa viidestä henkilöstä.
- Toiselle paikalle voi asettua joku jäljelle jääneestä neljästä henkilöstä.
- Kolmannelle paikalle voi asettua yksi kolmesta henkilöstä.
- Neljännelle paikalle on jäljellä enää kaksi vaihtoehtoa.
- Jonon viimeiselle paikalle on jäänyt enää yksi henkilö.



Kokonaisuus eli **jono** muodostuu viiden peräkkäisen vaiheen kautta. Jokaisessa vaiheessa suoritetaan tietty määrä valintoja. Jokaisen valinnan jälkeen seuraavassa vaiheessa erilaisten vaihtoehtojen määrä on yhtä pienempi.

Tuloperiaatteen mukaan erilaisia jonoja on $5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$. Jonojen määrän ilmaiseva lauseke voidaan merkitä lyhyesti **kertoman** avulla:

$$5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5!$$

Luetaan ”viiden kertoma”.

Kertomalla tarkoitetaan tuloa, jossa tietystä luvusta (esimerkiksi 5) alkaen peräkkäiset kokonaisluvut kerrotaan keskenään lukuun yksi asti. Kertoma on lyhyt ja helppo tapa esittää erilaisten jonojen määrä, kun jonon muodostavia alkioita on paljon. Esimerkiksi kymmenen alkion joukosta voidaan muodostaa erilaisia jonoja $\mathbf{10 \cdot 9 \cdot 8 \cdot \ldots \cdot 3 \cdot 2 \cdot 1 = 10!}$ kappaletta.

Laskimella: $x!$

Luvun n kertoma

n alkiota voidaan järjestää jonoon $n!$ erilaisella tavalla.

$n! = n \cdot (n-1) \cdot \ldots \cdot 3 \cdot 2 \cdot 1$

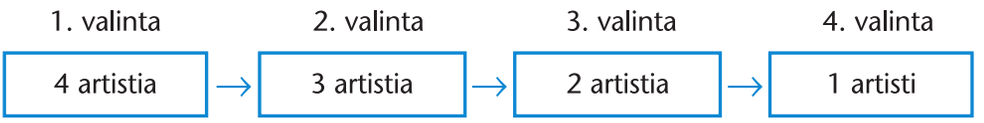
ESIMERKKI 2

Emma osallistuu tietokilpailuun, jonka ensimmäisenä tehtävänä on asettaa seuraavat artistit ikäjärjestykseen nuorimmasta alkaen: Pelle Miljoona, Martti Syrjä, Hector ja Neuman. Emma ei tiedä artistien ikiä, joten hän arvaa järjestyksen. Millä todennäköisyydellä hän laittaa artistit oikeaan järjestykseen?

RATKAISU

Ensimmäinen artisti voidaan valita neljän artistin joukosta. Koska jo valittua artistia ei voi enää valita, toinen artisti voidaan valita jäljelle jääneen kolmen joukosta. Kolmannessa valinnassa on jäljellä vain kaksi ja viimeisessä vain yksi artisti.

Tuloperiaatteen mukaan erilaisia ikäjärjestyksiä on 4! = 24.



Oikeita ikäjärjestyksiä on vain yksi, joten $P(\text{oikea järjestys}) = \frac{1}{24}$.

VASTAUS $\frac{1}{24}$

ESIMERKKI 3

Pohjoismaissa toimivan yrityksen kokoukseen osallistui yhdeksän henkilöä, joista kolme oli suomalaisia ja kaksi ruotsalaisia. He sopivat, että kukin hoitaa tiedotusta neljä kuukautta, jonka jälkeen vastuu siitä siirtyy seuraavalle. Järjestys suoritettiin arpomalla. Kuinka monta sellaista järjestystä on, jossa

a) suomalaiset hoitavat kolme ensimmäistä vuoroa
b) ruotsalaiset hoitavat alun kaksi vuoroa ja suomalaiset kolme viimeistä vuoroa?

RATKAISU

a) Suomalaiset voidaan järjestää 3! = 6 eri tavalla. Sen jälkeen loput henkilöt voidaan järjestää 6! = 720 eri tavalla. Tuloperiaatteen mukaan erilaisia järjestyksiä, joissa suomalaiset ovat alussa, on

$$3! \cdot 6! = 6 \cdot 720 = 4320$$

b) Ruotsalaiset voidaan järjestää alkuun 2! = 2 eri tavalla. Tämän jälkeen tulevat muut pohjoismaalaiset paitsi suomalaiset. Heidät voidaan järjestää 4! eri tavalla. Suomalaiset voidaan järjestää loppuun 3! tavalla.

Erilaisia järjestyksiä on siis yhteensä $2 \cdot 4! \cdot 3! = 288$.

VASTAUS a) 4320 b) 288

Kombinaatiot

Kun perusjoukosta muodostetaan pienempiä osajoukkoja, saadaan erilaisia **kombinaatioita**. Osajoukossa alkioiden järjestyksellä ei ole merkitystä. Kaksi osajoukkoa ovat erilaiset, jos niissä on yksikin erilainen alkio.

Tarkastellaan esimerkkinä viiden henkilön joukkoa A, B, C, D ja E. Muodostetaan tästä perusjoukosta kolmen henkilön osajoukkoja, jotka ovat erilaisia:

ABC	ABD	ABE	ACD	ACE
ADE	BCD	BCE	BDE	CDE

Joukkoja on kymmenen erilaista. Esimerkiksi joukko {ACB} ei kelpaa muodostettujen kombinaatioiden joukkoon, koska sen alkiot ovat samat kuin joukolla {ABC}.

Kun henkilöiden määrä kasvaa, erilaisten kolmen henkilön osajoukkojen lukumäärä kasvaa nopeasti. Esimerkiksi kymmenen henkilön perusjoukosta voidaan muodostaa jo 120 erilaista kolmen henkilön osajoukkoa. Niinpä kombinaatioiden etsiminen taulukoimalla ei ole järkevää. Määritellään laskutoimitus, joka laskee erilaisten kombinaatioiden määrän.

Jos kymmenen alkion perusjoukosta muodostetaan kolmen alkion osajoukkoja, niitä on yhteensä

$$\binom{10}{3} = 120$$

Luetaan ”kymmenen yli kolmen”.

Useimmissa laskimissa kombinaatiot saadaan näppäimellä nCr. Kombinaatio voidaan laskea myös kertoman avulla.

Kombinaatio

n:n alkion perusjoukosta valitaan osajoukkoja, joiden koko on k

$$\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$$

ESIMERKKI 4 Korttipakasta nostetaan kuusi korttia palauttamatta kortteja pakkaan.

a) Kuinka monta erilaista kuuden kortin kättä voidaan muodostaa?

b) Kuinka monta erilaista kuusi herttaa sisältävää kättä voidaan muodostaa?

c) Käteen halutaan neljä herttaa. Kuinka monta erilaista neljä herttaa sisältävää kättä voidaan muodostaa?

RATKAISU

a) Korttipakan kortit, joita on 52 kappaletta, muodostavat perusjoukon. Tästä joukosta on muodostettava kuuden kortin osajoukkoja.

Erilaisia kuuden kortin käsiä on $\binom{52}{6} = 20\,358\,520$.

Laskimella näppäillään 52 **nCr** 6

b) Perusjoukko muodostuu nyt pelkistä hertoista, koska kaikki kuusi korttia ovat herttoja. Yhteensä pakassa on 13 herttaa.

Erilaisia kuusi herttaa sisältäviä käsiä on $\binom{13}{6} = 1716$.

c) Neljä herttaa valitaan 13 hertan joukosta. Loput kaksi korttia voidaan valita jäljelle jääneen 39 muun kortin joukosta.

Erilaisia neljän hertan yhdistelmiä on $\binom{13}{4} = 715$.

Erilaisia kahden ei-hertan yhdistelmiä on $\binom{39}{2} = 741$.

Tuloperiaatteen mukaan haluttuja kuuden kortin käsiä on

$$\binom{13}{4} \cdot \binom{39}{2} = 715 \cdot 741 = 529\,815$$

VASTAUS a) 20 358 520 b) 1716 c) 529 815

ESIMERKKI 5 Belgialainen lotto sisältää 45 numeroa, joista arvotaan kuusi voittonumeroa. Millä todennäköisyydellä pelaaja saa lottoarvonnassa

a) kaikki kuusi oikein

b) täsmälleen neljä oikein?

RATKAISU a) Erilaisia lottorivejä on yhteensä $\binom{45}{6}$ kappaletta.

Oikeat numerot täytyy valita lottoarvonnan tuloksen kuuden oikean numeron joukosta.

Rivejä, joissa on kaikki kuusi numeroa oikein, on $\binom{6}{6} = 1$

$$P(\text{kaikki kuusi oikein}) = \frac{1}{\binom{45}{6}} = \frac{1}{8\,145\,060} \approx 1{,}23 \cdot 10^{-7}$$

b) Neljä oikeaa numeroa täytyy valita kuuden voittonumeron joukosta. Loput kaksi numeroa voidaan valita $45 - 6 = 39$ väärän numeron joukosta.

Neljä oikeaa numeroa voidaan valita $\binom{6}{4} = 15$ eri tavalla.

Kaksi väärää numeroa voidaan valita $\binom{39}{2} = 741$ eri tavalla.

Tuloperiaatteen mukaan erilaisia neljä oikein -rivejä on siis $\binom{6}{4} \cdot \binom{39}{2}$.

$$P(\text{täsmälleen neljä oikein}) = \frac{\binom{6}{4} \cdot \binom{39}{2}}{\binom{45}{6}} = \frac{15 \cdot 741}{8\,145\,060} \approx 1{,}36 \cdot 10^{-3}$$

VASTAUS a) $1{,}23 \cdot 10^{-7}$ b) $1{,}36 \cdot 10^{-3}$

Tehtäviä

Tulo-periaate

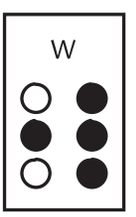
173. Jaakon kaapissa on neljät housut, viisi paitaa ja kolmet kengät. Kuinka monta erilaista asukokonaisuutta niistä voidaan muodostaa?

174. Auton rekisterikilven alkuosa koostuu yleensä kolmesta kirjaimesta ja loppuosa kolmesta numerosta. Kuinka monta erilaista rekisterikilpeä tällä periaatteella voidaan muodostaa?

E1 175. Henkilötunnuksen loppuosa koostuu yksilönumerosta eli kolmen numeron sarjasta sekä sen perässä olevasta tarkistusmerkistä. Tarkistusmerkki voi olla joko numero tai kirjain. Kuinka monta sellaista tunnusta voidaan muodostaa, jossa tarkistusmerkkinä on kirjain? Kirjaimia on käytössä 21.

176. Kuuntelukokeen kymmenestä kohdasta jokainen sisältää neljä vaihtoehtoa.
   a) Kuinka monta erilaista vastausriviä on olemassa?
   b) Millä todennäköisyydellä opiskelija saa arvaamalla kaikki kohdat oikein?

177. Ranskalaisen *Louis Braillen* vuonna 1825 kehittämä pistekirjoitus on kohokirjoitusta, jota luetaan sormin. Pistekirjoitusjärjestelmässä kutakin merkkiä kohti on käytettävissä kuusi kiinteää paikkaa, joihin voidaan asettaa yhdestä kuuteen pistettä. (Esimerkkinä kuviossa on kirjain W.) Kuinka monta erilaista merkkiä järjestelmässä voidaan esittää? [K2007, 10]



178. Kulhossa on viisi punaista ja kymmenen mustaa palloa. Kulhosta poimitaan umpimähkään viisi palloa palauttamatta yhtäkään. Mikä on todennäköisyys, että ainakin yksi poimituista palloista on punainen. Millä todennäköisyydellä kaikki viisi ovat samanvärisiä? [K2008, 8]

179. Mäkihypyn toiselle kierrokselle on selvinnyt 30 kilpailijaa. Kuinka monta erilaista hyppyjärjestystä heistä voitaisiin muodostaa?

E2 180. ”*Sound of music*” on tunnettu musikaali, joka kertoo itävaltalaisesta von Trappin perheestä. Kotiopettajatar Maria asettaa perheen seitsemän lasta jonoon arpomalla. Millä todennäköisyydellä lapset tulevat jonoon ikäjärjestyksessä joko vanhimmasta nuorimpaan tai nuorimmasta vanhimpaan?

E3 181. Kolme poikaa ja kolme tyttöä asettuu jonoon siten, että pojat ovat jonon loppupäässä. Kuinka monta erilaista jonoa voidaan muodostaa?

182. Kuudesta henkilöstä, joista kaksi on naisia, muodostetaan umpimähkään jono. Millä todennäköisyydellä naiset ovat jonossa peräkkäin?

183. Maantietoon perehtymätön henkilö saa eteensä viiden valtion nimet. Hänen on asetettava maat suuruusjärjestykseen pienimmästä alkaen. Millä todennäköisyydellä kaksi pienintä maata ovat oikeilla paikoillaan, mutta lopuista maista ainakin kaksi on väärällä paikalla?

Kombinaatiot

184. Kirjoita ja laske lausekkeet kertoman avulla ja tarkasta sitten laskimella.

a) $\binom{7}{3}$ b) $\binom{13}{6}$

185. Mäkihyppykilpailun tuomaristossa on viisi varsinaista jäsentä. Jäsenet valitaan 27 osallistujamaan joukosta. Jos jokaisesta maasta valittaisiin enintään yksi tuomari ja valinta suoritettaisiin arpomalla, kuinka monta erilaista tuomaristoa voitaisiin muodostaa?

186. Päivällisille osallistuu 15 henkilöä. Ruokailijat sijoitetaan kolmen henkilön pöytiin ja istumapaikat arvotaan. Kuinka monta erilaista kolmen ruokailijan kokoonpanoa voidaan muodostaa?

E4 187. Useissa korttipeleissä pelaajalle jaetaan viisi korttia. Kuinka monta
a) viiden kortin kättä voidaan muodostaa
b) sellaista kättä on olemassa, joissa kaikki kortit ovat punaisia maita (eli herttoja tai ruutuja)
c) sellaista kättä on olemassa, joissa on neljä ässää?

188. Tietokilpailun kuuden parhaan kesken arvotaan kolme lahjakorttia. Millä todennäköisyydellä lahjakortin saavat tähän joukkoon sijoittuneet Sara, Mimmi ja Harri?

189. Kutsuilla on 10 avioparia. Jokainen läsnäolija kättelee jokaista muuta paitsi aviopuolisoaan. Kuinka monta kättelyä suoritetaan? [K1998, 3b]

190. Luokan pojista kymmenen on yli 180 cm pitkiä ja kahdeksan alle 180 cm pitkiä. Pojista valitaan umpimähkään neljä. Millä todennäköisyydellä kaksi heistä on yli ja kaksi alle 180 cm?

E5 191. Lottoarvonnassa arvotaan 39 numerosta 7 numeroa. Kuinka monta erilaista 7 rastin lottoruudukkoa on olemassa? Kun on arvottu seitsemän numeroa (eikä ns. lisänumeroita oteta huomioon), erilaisia 4 oikein -ruudukkoja on erään lehtiartikkelin mukaan olemassa 173 600. Tarkista, onko luku oikea. Jos ei, niin mikä on oikea lukumäärä? Laske, kuinka monta erilaista 5 oikein -ruudukkoa ja kuinka monta erilaista 6 oikein -ruudukkoa on olemassa. [S2003, 9]

192. Sienikurssilla opetettiin tunnistamaan 78 erilaista sientä, joista kurssilainen oppi kuitenkin vain 49. Kuinka suurella todennäköisyydellä hän tunnisti oikein satunnaisesti esitetyt kuusi erilaista kurssilla opetettua sientä? [S1996, 7c]

193. Laatikossa on 10 sinistä ja 7 punaista palloa. Laatikosta nostetaan umpimähkään 6 palloa. Millä todennäköisyydellä
a) yksi palloista on sininen
b) kolme palloista on punaisia?

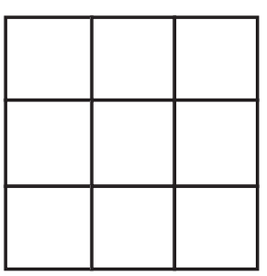
Sekalaisia tehtäviä

194. Kaisa pakkaa matkaa varten vaatteita. Hänen vaatekaapissaan on kuudet housut ja kahdeksan paitaa. Hän arpoo matkalle mukaan kolmet housut ja kolme paitaa.
   a) Kuinka monta erilaista vaatekokonaisuutta Kaisa voi matkalla muodostaa?
   b) Jos Kaisa ei valitse vaatteita arpomalla, kuinka monella tavalla hän voi valita matkalle otettavat vaatteet?

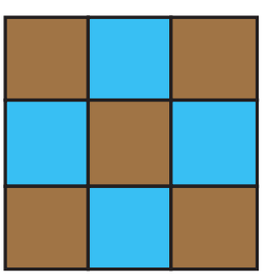
195. Ryhmässä on kolme blondia, kaksi brunettea ja kaksi tummaverikköä.
   a) Kuinka monta erilaista jonoa heistä voi muodostaa?
   b) Kuinka monta sellaista jonoa on olemassa, jossa blondit ovat jonon ensimmäisinä?
   c) Kuinka monta sellaista jonoa on olemassa, jossa brunette on jonon ensimmäisenä ja viimeisenä?
   d) Kuinka monta sellaista jonoa on olemassa, jossa blondit ovat peräkkäin?

196. Luokassa on 28 oppilasta. Kuinka monella tavalla oppilaiden joukosta voidaan valita
   a) kolmihenkinen toimikunta
   b) viisihenkinen toimikunta?

197. Alla olevan kuvion 1 kukin ruutu väritetään satunnaisesti ja toisista riippumatta joko ruskeaksi tai siniseksi.
   a) Millä todennäköisyydellä saadaan kuvion 2 shakkilautakuvio?
   b) Millä todennäköisyydellä mikään vaakarivi ei ole yksivärinen? (K2011, 8)



Kuvio 1



Kuvio 2

198. Laatikossa on 17 tuoretta ja 5 pilaantunutta tomaattia. Laatikosta nostetaan umpimähkään viisi tomaattia. Millä todennäköisyydellä
   a) kaikki nostetut tomaatit ovat tuoreita
   b) vain yksi tomaatti on tuore?

2.5 Binomitodennäköisyys

POHDINTA 1 Tapahtuma ”heitetään viidellä nopalla kolme ykköstä” sisältää useita eri tilanteita, esimerkiksi seuraavat:

a) Laske edellä esitetyn kolmen tapauksen todennäköisyydet.

b) Kuinka monta erilaista suotuisaa heittosarjaa tapahtuma ”heitetään viidellä nopalla kolme ykköstä” sisältää?

c) Laske a- ja b-kohtien avulla todennäköisyys P(heitetään viidellä nopalla kolme ykköstä).

Tapahtumaa, jossa samaa koetta toistetaan useita kertoja, kutsutaan **toistokokeeksi**. Esimerkiksi koripallossa koriinheitot ovat saman tapahtuman toistoa. Koriinheitto on siis toistokoe. Heitot ovat toisistaan riippumattomia. Tarkasteltava tapahtuma voi jokaisella toistokerralla joko onnistua tai epäonnistua, muita vaihtoehtoja ei ole.

Tarkastellaan koripallossa tilannetta, jossa neljästä vapaaheitosta kaksi tulisi saada koriin. Merkitään onnistunutta heittoa kirjaimella **O** ja epäonnistunutta heittoa kirjaimella **E**.

- Oletetaan, että heitto onnistuu todennäköisyydellä
  $P(\mathbf{O}) = 0{,}75$
- Heitto epäonnistuu tällöin todennäköisyydellä
  $P(\mathbf{E}) = 1 - P(\mathbf{O}) = 1 - 0{,}75 = 0{,}25$
- Yhden suotuisan heittosarjan, jossa kaksi heittoa neljästä onnistuu, todennäköisyys on
  $P(\mathbf{OOEE}) = 0{,}75 \cdot 0{,}75 \cdot 0{,}25 \cdot 0{,}25 = 0{,}75^2 \cdot 0{,}25^2$

Tämä on kuitenkin vain yksi mahdollinen heittosarja. Erilaisia suotuisia heittosarjoja on:

OOEE
EOEO
EOOE
OEOE
EEOO
OEEO

Erilaisten heittosarjojen lukumäärä voidaan laskea kombinaation avulla: $\binom{4}{2} = 6$

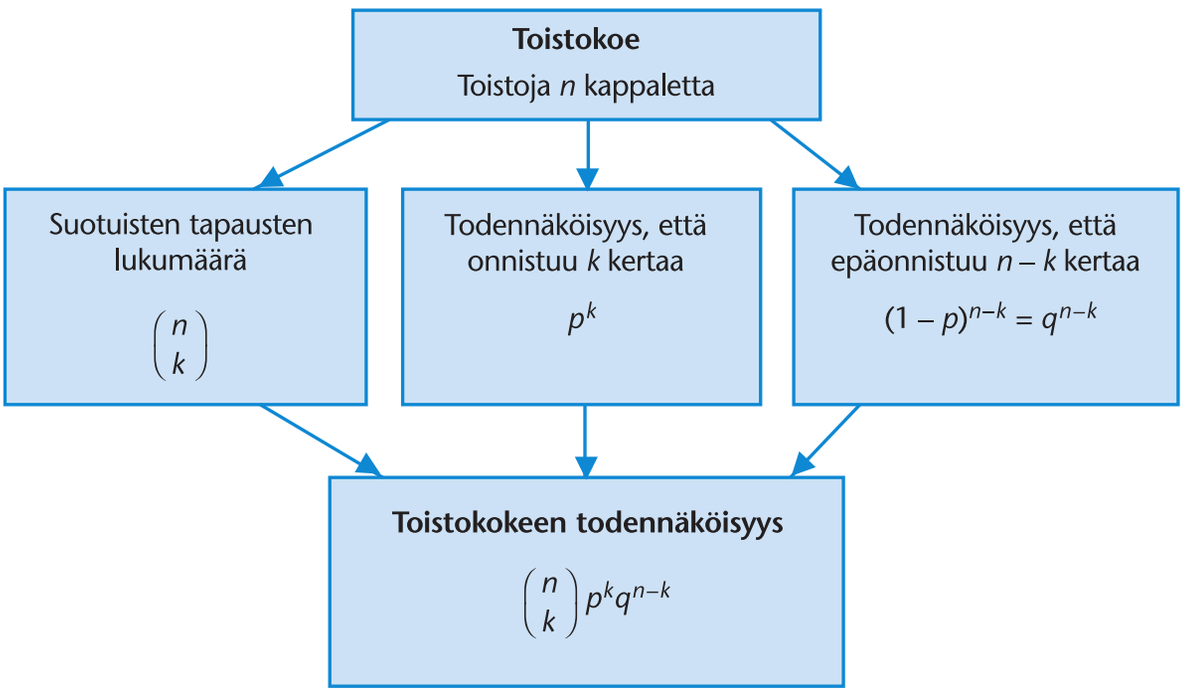
Todennäköisyys, että heitto onnistuu kahdesti, saadaan yhteenlaskusäännön avulla laskemalla erilaisten suotuisten heittosarjojen todennäköisyydet yhteen.

$$
\begin{aligned}
&P(\text{heitto onnistuu kahdesti}) \\
&= P(\mathbf{OOEE}) + P(\mathbf{OEOE}) + P(\mathbf{EOEO}) \\
&\quad + P(\mathbf{EEOO}) + P(\mathbf{EOOE}) + P(\mathbf{OEEO}) \\
&= 6 \cdot 0{,}75^2 \cdot 0{,}25^2 \\
&= 0{,}2109\ldots \\
&\approx 0{,}21
\end{aligned}
$$

Jos samaa tapahtumaa toistetaan useita kertoja, ei ole mielekästä luetella suotuisten kokeiden sarjoja, vaan suotuisten tapausten lukumäärä lasketaan kombinaatioilla.

Jos samaa koetta toistetaan n kertaa ja se onnistuu k kertaa, suotuisia tapahtumia on $\binom{n}{k}$ kappaletta.

Merkitään kokeessa onnistumisen todennäköisyyttä kirjaimella p. Kokeessa epäonnistumisen todennäköisyys on tällöin $q = 1 - p$.



Toistokokeen todennäköisyys saadaan siis laskemalla yhteen kaikkien suotuisten tapausten todennäköisyydet. Toistokokeen todennäköisyyttä kutsutaan myös **binomitodennäköisyydeksi**.

Binomitodennäköisyys

$$\binom{n}{k} p^k q^{n-k}$$

- n = toistojen lukumäärä
- k = onnistumisten lukumäärä
- p = onnistumisen todennäköisyys
- q = epäonnistumisen todennäköisyys

ESIMERKKI 1 Vakioveikkauksessa on kolmetoista kohdetta. Jokaisessa kohteessa on kolme mahdollisuutta (1, X, 2). Veikkaaja arvaa jokaisen kohteen umpimähkään. Laske todennäköisyys, että saadaan täsmälleen kymmenen oikein.

RATKAISU Määritetään ensin onnistumisen ja epäonnistumisen todennäköisyydet:

- Todennäköisyys, että veikataan yksi kohde oikein: $p = \frac{1}{3}$.
- Todennäköisyys, että veikataan yksi kohde väärin: $q = 1 - \frac{1}{3} = \frac{2}{3}$.

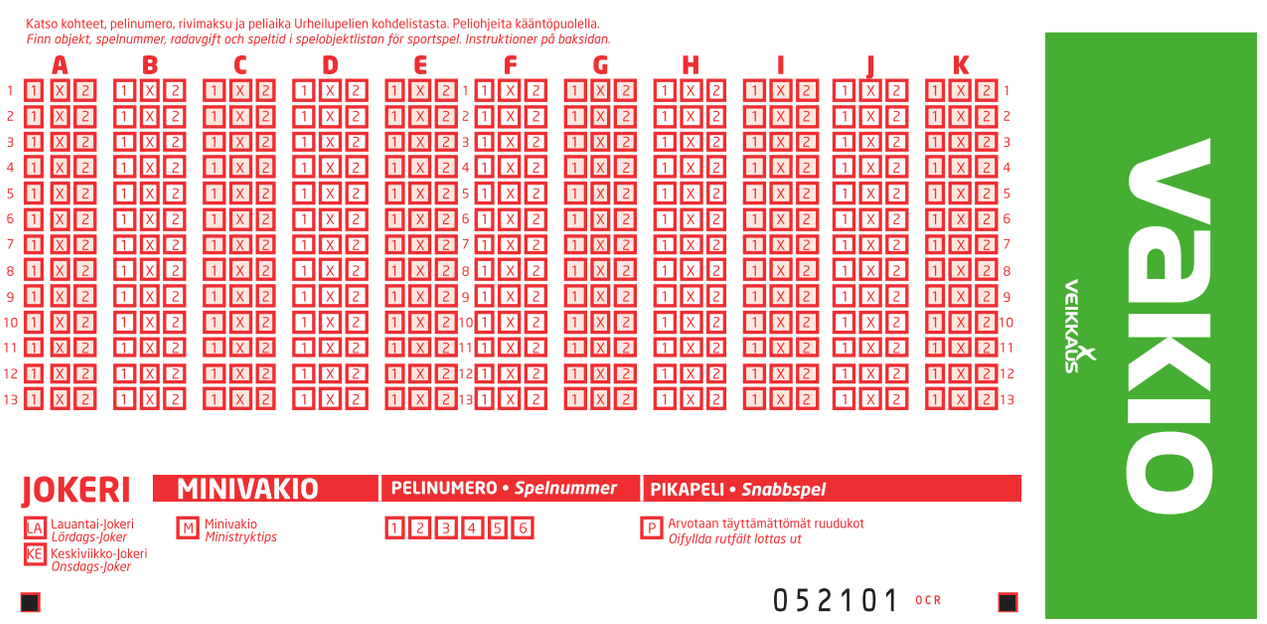
Kolmestatoista kohteesta voidaan valita kymmenen $\binom{13}{10}$ eri tavalla.

Todennäköisyys saada kymmenen oikein on

$$\binom{n}{k} p^k q^{n-k}$$

$$\begin{aligned} P(\text{kymmenen oikein}) &= \binom{13}{10}\left(\frac{1}{3}\right)^{10}\left(\frac{2}{3}\right)^{13-10} \\ &= 1{,}435\ldots \cdot 10^{-3} \\ &\approx 0{,}0014 \end{aligned}$$

VASTAUS Todennäköisyys saada täsmälleen kymmenen oikein on 0,0014.



ESIMERKKI 2 Alueen merikotkapopulaatiosta 28 % on rengastettuja. Seurantatutkimuksissa tarkastettiin 12 lintua satunnaisotannalla. Millä todennäköisyydellä

a) korkeintaan kaksi linnuista oli rengastettu

b) ainakin kaksi linnuista oli rengastettu?

RATKAISU Koska merikotkapopulaatiosta 28 % on rengastettuja, todennäköisyys, että satunnaisesti valittu kotka on rengastettu, on $p = 0{,}28$.

Todennäköisyys, että kotkaa ei ole rengastettu, on $q = 1 - 0{,}28 = 0{,}72$

a) Tapahtuma ”korkeintaan kaksi linnuista on rengastettu” sisältää tapahtumat

- A = ”yhtään lintua ei ole rengastettu”
- B = ”yksi linnuista on rengastettu”
- C = ”kaksi linnuista on rengastettu”

$\binom{n}{k} p^k q^{n-k}$

$$\begin{aligned}
&P(\text{korkeintaan kaksi linnuista on rengastettu}) \\
&= P(A) + P(B) + P(C) \\
&= \binom{12}{0} \cdot 0{,}28^0 \cdot 0{,}72^{12} + \binom{12}{1} \cdot 0{,}28^1 \cdot 0{,}72^{11} + \binom{12}{2} \cdot 0{,}28^2 \cdot 0{,}72^{10} \\
&= 0{,}3037... \\
&\approx 0{,}30
\end{aligned}$$

b) Tapahtuma ”ainakin kaksi linnuista on rengastettu” on hankala laskea suoraan, joten käytetään komplementtia ”korkeintaan yksi linnuista on rengastettu”. Komplementti sisältää tapahtumat A ja B.

$$\begin{aligned}
&P(\text{ainakin kaksi linnuista on rengastettu}) \\
&= 1 - P(\text{korkeintaan yksi linnuista on rengastettu}) \\
&= 1 - [P(A) + P(B)] \\
&= 1 - \left[\binom{12}{0} \cdot 0{,}28^0 \cdot 0{,}72^{12} + \binom{12}{1} \cdot 0{,}28^1 \cdot 0{,}72^{11} \right] \\
&= 1 - 0{,}1099... \\
&= 0{,}8900... \\
&\approx 0{,}89
\end{aligned}$$

VASTAUS a) 0,30 b) 0,89

Tehtäviä

199. Laske laskimella seuraavien lausekkeiden arvot.

a) $\binom{7}{3} \cdot 0{,}25^3 \cdot 0{,}75^4$ b) $\binom{9}{2} \cdot \left(\frac{2}{7}\right)^2 \cdot \left(\frac{5}{7}\right)^5$

200. Erään tapahtuman todennäköisyys saatiin laskettua lausekkeella $\binom{12}{5} \cdot 0{,}90^5 \cdot 0{,}10^7$. Laadi tähän ratkaisuun sopiva ongelma.

E1 201. Oona pelaa pesäpalloa. Hän osuu lyönnillään palloon 70 prosentin todennäköisyydellä. Oona lyö pelissä 24 lyöntiä. Millä todennäköisyydellä Oona osuu palloon täsmälleen 15 kertaa?

202. Erään koirarodun yksilöistä 35 % on valkoruskeita ja loput mustavalkoisia. Millä todennäköisyydellä seitsemän pennun pentueessa on
a) kaksi mustavalkoista pentua
b) neljä valkoruskeaa pentua?

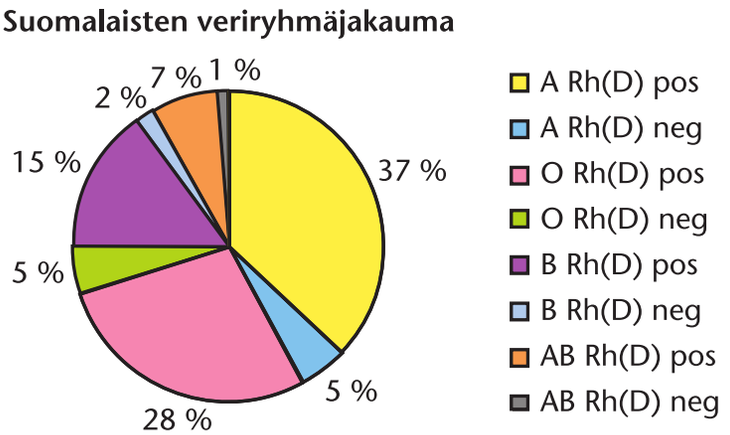
203. Nurmikonsiemenistä 88 % itää. Laske todennäköisyys, että 50 siemenen pussissa
a) on täsmälleen 45 itävää siementä
b) kaikki siemenet ovat itävät.

204. Koulun käytävällä tehdään gallupkyselyä, johon vastaajiksi valitaan sattumanvaraisesti 100 oppilasta. Tytöillä ja pojilla on yhtä suuri mahdollisuus osallistua kyselyyn. Millä todennäköisyydellä vastaajina on täsmälleen yhtä monta tyttöä ja poikaa?

205. Suomalaisista koululaisista on erään tilaston mukaan 5 % vasenkätisiä. Laske sen tapahtuman todennäköisyys, että 32 oppilaan luokassa on 4 vasenkätistä. [K1995, 9]

E2 206. Vaateliikkeen myymistä vaatteista 32 % on valmistettu Suomessa. Asiakas valitse satunnaisesti sovitettavakseen 9 vaatetta. Millä todennäköisyydellä vaatteista
a) korkeintaan kaksi on valmistettu Suomessa
b) vähintään kaksi on valmistettu Suomessa?

207. Terveydenhoitaja valitsee luokan opiskelijoista satunnaisesti 10 henkilöä terveystarkastukseen. Millä todennäköisyydellä tarkastukseen valituista opiskelijoista
a) neljä kuuluu veriryhmään A
b) vähintään kaksi kuuluu veriryhmään A
c) korkeintaan kaksi kuuluu veriryhmään A?
Eri veriryhmät esiintyvät seuraavat jakauman mukaisesti.



Lähde: Punainen Risti, veripalvelu

208. Makeistehtaan valmistamista suklaalevyistä 3 % on viallisia. Laske todennäköisyys, että satunnaisesti valitussa kymmenen levyn erässä on
a) yksi viallinen levy
b) ainakin kolme viallista levyä.

209. Kun hyönteiskanta on suurimmillaan, islanninponin todennäköisyys sairastua kesäihottumaan on 16,7 %. Tallilla on 14 islanninponia. Millä todennäköisyydellä kesäihottumaa saa
a) viisi ponia
b) ainakin kolme ponia?
Oletetaan ponin sairastumisen olevan riippumatonta muiden ponien sairastumisesta.

210. Erään liikkeen CD-levyistä 96 % on virheettömiä. Ostaja tarkastaa 21 levyn erän. Millä todennäköisyydellä hän löytää
a) kaksi virheellistä levyä
b) ainakin kaksi virheellistä levyä?

211. Noppaa heitetään 5 kertaa. Millä todennäköisyydellä tuloksena on
a) täsmälleen kaksi kuutosta
b) vähintään kaksi kuutosta?
[K2006, 9]

Sekalaisia tehtäviä

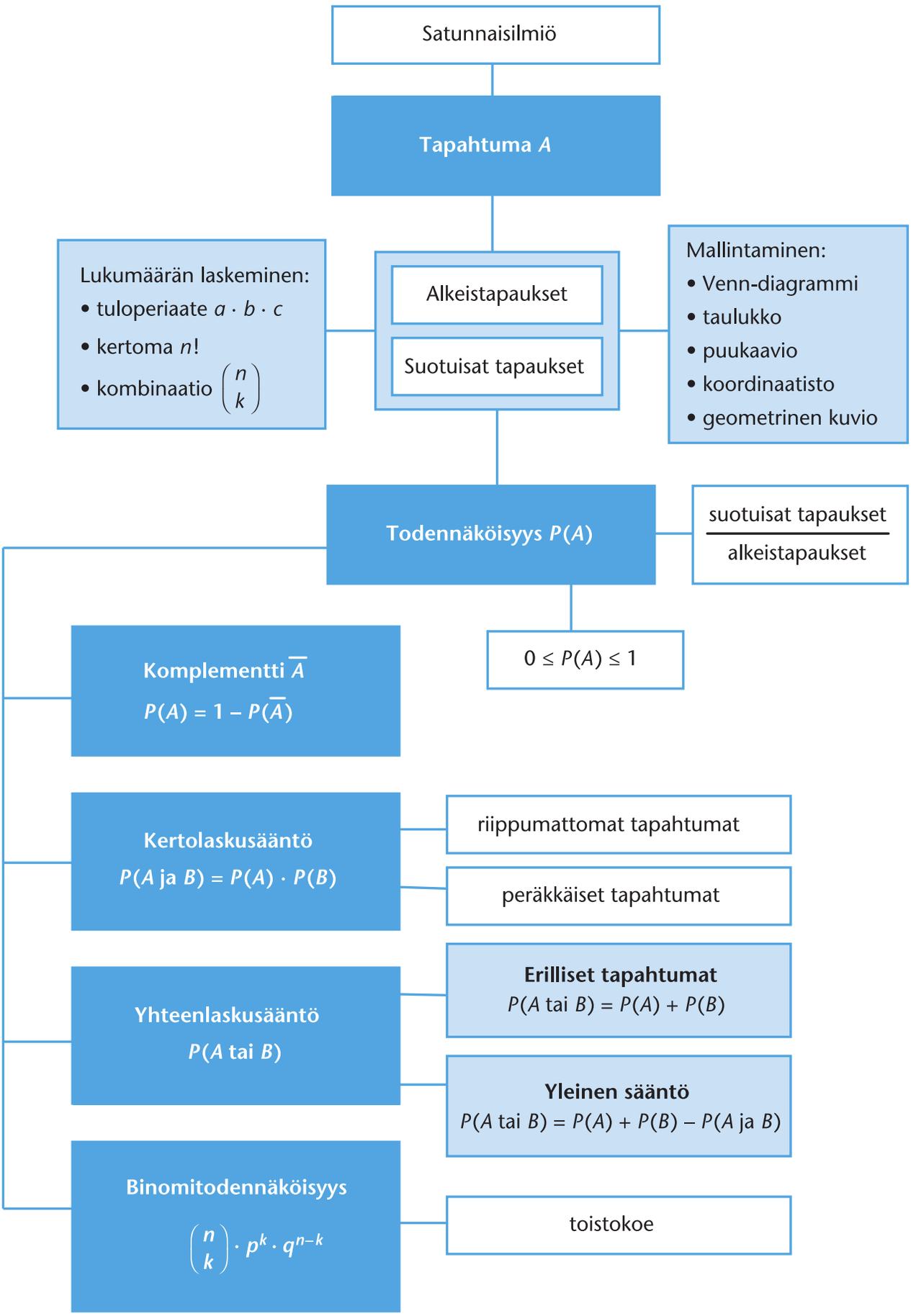
212. Noppaa heitetään kuusi kertaa. Millä todennäköisyydellä saadaan kahdesti kuutonen?

213. Pekka tietää kokemuksesta, että hän osuu tikanheitossa kymppiin 8 % todennäköisyydellä. Pekka heittää viisi tikkaa. Millä todennäköisyydellä hän osuu kymppiin
   a) 1. mutta ei muilla tikoilla
   b) ainakin yhdellä tikalla
   c) täsmälleen yhdellä tikalla? [K2000, 6]

214. Ihmisten syntymäajat tallennetaan erääseen tietokoneeseen muodossa PPKKVV. Tallentaja painaa väärää nappia 4,9 %:n todennäköisyydellä. Millä todennäköisyydellä kuusinumeroiseen koodiin tulee
   a) yksi virhe
   b) enintään yksi virhe?

215. Vaaleasilmäisiksi kutsutaan henkilöitä, joilla on siniset, harmaat tai vihreät silmät. Suomalaisista 89 % on vaaleasilmäisiä. Opiskelijaryhmästä valitaan satunnaisesti 8 opiskelijaa. Millä todennäköisyydellä
   a) ainakin kuusi opiskelijaa on vaaleasilmäisiä
   b) korkeintaan kaksi opiskelijaa on vaaleasilmäisiä?

Keskeisiä käsitteitä



3 Kertausosa

1.1 Tilastotieteen peruskäsitteitä

- Tutkimusta varten kerättyä aineistoa kutsutaan **havaintoaineistoksi**. Havaintoaineisto koostuu **havaintoyksiköistä**.
- **Muuttuja** on havaintoyksikköön liittyvä tutkittava ominaisuus.
- **Perusjoukko** eli **populaatio** koostuu kaikista havaintoyksiköistä, joista tietoa halutaan kerätä.
- **Otos** on perusjoukon osa, joka on kerätty jollain otantamenetelmällä.
- **Frekvenssi** f kertoo, kuinka monta kertaa muuttujan arvo esiintyy havaintoaineistossa.
- **Frekvenssijakauma** on taulukko, jossa on muuttujan arvot ja niihin liittyvät frekvenssit.
- **Suhteellinen frekvenssi** f % kertoo esiintymiskertojen prosentuaalisen osuuden havaintoyksiköiden kokonaismäärästä.
- **Moodi Mo** on se muuttujan arvo, jolla on suurin frekvenssi. Luokitellussa aineistossa moodina voidaan käyttää luokan luokkakeskusta.

1. Lukion lyhyen matematiikan äänimäärät erään lukion kevään ylioppilaskirjoituksissa olivat seuraavat:

Äänimäärä	0	2	3	4	5	6	7
Kpl	1	4	7	15	18	12	2

   a) Laadi aineistosta absoluuttinen ja suhteellinen frekvenssijakauma.
   b) Määritä aineiston moodi.
   c) Havainnollista suhteellista frekvenssijakaumaa sektoridiagrammilla.

2. Alla olevassa taulukossa on erään lukion opiskelijoiden preliminäärikokeeseen käyttämä aika.

Aika (min)	Opiskelijoita
0–59	2
60–119	3
120–179	7
180–239	28
240–299	25
300–359	11

   a) Muodosta absoluuttinen ja suhteellinen frekvenssijakauma.
   b) Määritä aineiston moodi.
   c) Havainnollista absoluuttista frekvenssijakaumaa kuvaajan avulla.

3. Alla on lueteltu erään matematiikan kurssin opiskelijoiden pituuksia (cm). Luokittele aineisto sopivasti ja määritä luokkakeskukset. Muodosta absoluuttinen ja suhteellinen frekvenssijakauma.

159	165	167	166	170	171	169
175	173	168	167	180	184	175
170	169	169	171	168	180	183

- Frekvenssin ja suhteellisen frekvenssin kuvaaminen
  - Epäjatkuva eli diskreetti muuttuja: pylväsdiagrammi, palkkidiagrammi, sektoridiagrammi
  - Jatkuva muuttuja: histogrammi
- luok.kesk. $= \dfrac{\text{tod. alaraja} + \text{tod. yläraja}}{2}$

4. Kemian kurssikokeen arvosanat jakautuivat seuraavasti:

Arvosana	4	5	6	7	8	9	10
Kpl	2	3	4	8	6	4	3

Laske todennäköisyys, että opiskelija saa arvosanaksi
a) 8 b) vähintään 7 c) enintään 6.

1.2 Summafrekvenssit

- **Summafrekvenssi *sf*** saadaan laskemalla frekvenssejä yhteen. Summefrekvenssi kertoo kuinka paljon havaintoja on kertynyt tiettyyn muuttujan arvoon mennessä.
- **Suhteellinen summafrekvenssi *sf* %** kuvaa kertymää prosentteina. Suhteellinen summafrekvenssi saadaan
  - laskemalla suhteelliset frekvenssit yhteen tai
  - ilmoittamalla summafrekvenssi prosentteina (tarkin tapa).
- Summafrekvenssin ja suhteellisen summafrekvenssin kuvaaminen
  - diskreetti muuttuja: porrasdiagrammi
  - jatkuva muuttuja: summakäyrä
- **Mediaani Md** kuuluu keskilukuihin. Se on suuruusjärjestykseen laitetun aineiston keskimmäinen arvo.

5. Kevään 2012 ylioppilastutkinnon lyhyen matematiikan kokeen arvosanajakauma oli seuraava:

Arvosana	I	A	B	C	M	E	L
Kpl	437	829	1863	2388	2021	1653	528

Lähde: YTL

a) Muodosta summafrekvenssijakauma ja suhteellinen summafrekvenssijakauma.
b) Kuinka monta prosenttia kokelaista sai arvosanakseen korkeintaan C?
c) Kuinka monta prosenttia kokelaista sai arvosanaksi vähintään M?

6. Vuonna 2011 Suomessa oli 264 622 20–44-vuotiasta henkilöä, joilla ei ollut perusasteen jälkeistä tutkintoa. Taulukkoon on koottu henkilöiden ikärakenne. Muodosta suhteellinen summafrekvenssijakauma ja määritä sen avulla mediaaniluokka.

Ikä	*f*
20–24	64 257
25–29	56 935
30–34	48 506
35–39	43 998
40–44	50 926

Lähde: Tilastokeskus

7. Tehdas valmistaa säilyketölkkejä. Laadunvalvonta tutki valmistettujen tölkkien tilavuuksia. Tulokset on koottu seuraavaan taulukkoon. Määritä suhteellisen summafrekvenssijakauman kuvaajan avulla säilyketölkin tilavuuden mediaani.

Tilavuus (cm^3)	*f*
190–194	240
195–199	760
200–204	780
205–209	320
210–214	250
215–219	170
220–224	140

1.3 Keskiarvo

- Sijaintiluvut jaetaan **keskilukuihin** ja **muihin sijaintilukuihin**.
- **Aritmeettinen keskiarvo** $\bar{x}$ kuvaa keskimääräistä muuttujan arvoa.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i x_i}{n}$$

- **Painotettu keskiarvo** lasketaan seuraavasti:
  1) Kerrotaan jokainen muuttujan arvo x_i omalla painokertoimellaan
  2) Lasketaan saadut tulot yhteen.
  3) Jaetaan summa painokertoimien summalla.

8. Luonnonsuojelualueella pesii runsaasti rastaita. Biologit kartoittivat rastaiden poikueitten kokoa pesintäaikaan laskemalla munat pesistä. Tulokset on koottu taulukkoon. Laske aineiston avulla rastaan pesän keskimääräinen poikueen koko.

Munia pesässä	Pesiä
1	3
2	4
3	15
4	12
5	33
6	27
7	8
8	2

9. Taulukossa on opintovelallisten 20–39-vuotiaiden ikärakenne Suomessa vuonna 2011. Määritä aineiston avulla opintovelallisten keski-ikä.

Ikä	20–24	25–29	30–34	35–39
Henkilöitä	82 660	98 111	63 251	27 451

Lähde: Tilastokeskus

10. Yrityksen A pörssikurssin kuusi viimeisintä noteerausta (€) olivat järjestyksessä:

14,15 13,00 12,80 12,61 11,81 12,12

Yritys B haluaa fuusioitua yrityksen A kanssa ja tarjoutuu ostamaan A:n osakkeet hinnalla, joka määräytyy seuraavasti: toiseksi viimeisin noteeraus saa painoarvon 2, viimeisin painoarvon 3 ja muut saavat painoarvon 1. Laske osakkeen hinta.

11. Lukujen $2x + 1$, x, $x - 3$ ja $-x + 5$ keskiarvo on 8. Mitkä luvut ovat?

12. Lämpömittareiden A ja B lukemia T_A ja T_B verrattiin tarkan lämpömittarin lukemiin T. Tulokset ovat oheisessa taulukossa. Laske mittareiden A ja B virheet $T_A - T$ ja $T_B - T$ ja vertaa mittareiden paremmuutta kussakin lämpötilassa T erikseen. Tutki, kumman mittarin virheiden keskiarvo on pienempi, ja vertaa tulosta aiempiin. Voidaanko mittareita paremmin vertailla jollakin muulla tunnusluvulla? [S1998, 7b]

T	+10,0	+6,0	+2,0	–2,0	–6,0	–10,0
T_A	+11,2	+7,1	+0,7	–1,2	–6,9	–10,7
T_B	+10,6	+6,5	+2,4	–1,6	–5,7	–9,8

1.4 Keskihajonta

- **Keskihajonta**

$$s_n = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

- **Otoskeskihajonta**

$$s_{n-1} = \sqrt{\frac{\sum_{i=1}^{k} f_i(x_i - \bar{x})^2}{n-1}}$$

- Jos muuttujan arvo on vähintään kahden keskihajonnan päässä keskiarvosta, se poikkeaa keskiarvosta merkitsevästi.

13. Erään kokeen arvosanojen frekvenssijakauma oli seuraava:

Arvosana	*f*
4	3
5	6
6	7
7	11
8	10
9	5
10	2

Laske kokeen keskiarvo ja keskihajonta.

14. Erään alueen metsänomistajien metsäpinta-alat olivat (ha):

13,4	23,6	45,3	21,1	34,6
70,2	89,4	110,3	46,9	98,7
85,4	12,5	54,9	67,3	87,5

a) Luokittele pinta-alat tasavälisiin luokkiin siten, että pienin luokka on 10–29.
b) Laske luokitellun aineiston keskiarvo ja keskihajonta.
c) Poikkeaako 89 ha:n metsänomistaja merkittävästi aineiston keskiarvosta?

15. Erään yrityksen työntekijöiden kuukausipalkat tilastoitiin.

Kuukausipalkka (€)	*f*
2500	2
3000	6
3340	7
3560	10
4010	6
4560	3
4930	1

a) Millä välillä palkat ovat, jos ne ovat enintään yhden hajonnan päässä keskiarvosta?
b) Poikkeaako 4560 €:n kuukausipalkka merkitsevästi tämän aineiston keskiarvosta?

1.5 Normaalijakauma

- Jatkuvan muuttujan jakaumista tärkein on normaalijakauma, jonka muoto riippuu keskiarvosta ja keskihajonnasta.

- Todennäköisyyslaskennassa normaalijakauman keskiarvoa vastaa **odotusarvo** μ. **Keskihajonnalle** käytetään merkintää σ.

- **Normitettu normaalijakauma**: $N(0, 1)$

- Muuttujan x **normitettu arvo** z

$$z = \frac{x - \bar{x}}{s} = \frac{x - \mu}{\sigma}$$

- Normitettuun normaalijakaumaan liittyvät todennäköisyydet saadaan kertymäfunktion taulukoitujen arvojen avulla.

16. Muuttuja x noudattaa normaalijakaumaa $N(75, 5)$. Laske todennäköisyys
a) $P(x < 70)$ b) $P(68 < x < 85)$.

17. Lukion A opiskelija Alina sai lukioiden yhteisessä vieraan kielen ainekirjoituskokeessa 82 pistettä ja lukion B opiskelija Bertta 80 pistettä. Kummassakin lukiossa kokeen keskiarvo oli 72 pistettä. Hajonta lukiossa A oli 9,2 pistettä ja lukiossa B 6,8 pistettä. Oletetaan, että pistemäärät noudattavat normaalijakaumaa. Tutki, kumpi opiskelija menestyi paremmin oman lukionsa tasoon verrattuna. Kuinka suuri osuus lukion B opiskelijoista menestyi kokeessa keskiarvoa paremmin, mutta huonommin kuin Bertta? Entä Alinaa paremmin lukiossa A? [S2001, 15]

18. Tehdas ilmoitti valmistamiensa auton sytytystulppien kestävyyden ajokilometreinä noudattavan normaalijakaumaa $N(25\,000 \text{ km}, 2000 \text{ km})$. Kuinka monen kilometrin jälkeen tällaisen tulpan toimintavarmuus laskee alle 95 prosentin? [S2002, 10]

19. Työhönottohaastattelu kestää keskimäärin 25 minuuttia. Oletetaan, että sen pituus on normaalisti jakaantunut $N(25, \sigma)$. Mikä on keskihajonnan σ oltava, että haastattelu ei 95 prosentin varmuudella ylitä 30:tä minuuttia? [S1999, 7a]

20. Sadan tulitikkurasian erästä 11 rasiaa sisältää vähintään 47 tulitikkua. Kuinka monta tulitikkua rasiassa on keskimäärin, kun tikkujen määrä rasiassa noudattaa likimain normaalijakaumaa ja keskihajonta on neljä tikkua?

2.1 Malleja todennäköisyyden määrittämiseksi

- Ilmiötä kutsutaan **satunnaisilmiöksi**, jos sattumalla on osuutensa tilanteessa.
- Satunnaisilmiöt sisältävät eri **tapahtumia**.
- Tapahtuman A **todennäköisyys** $0 \le P(A) \le 1$.
- **Alkeistapauksia** ovat kaikki mahdolliset tapahtumat, joita tilanteessa voi sattua.
- Alkeistapausten etsimiseen voi käyttää esimerkiksi
  – Venn-diagrammia
  – taulukkoa
  – koordinaatistoa.
- **Suotuisia tapauksia** ovat ne alkeistapaukset, jotka toteuttavat tapahtuman.
- Tapahtuman A **klassinen todennäköisyys**

$$P(A) = \frac{\text{suotuisien tapausten lkm.}}{\text{kaikkien alkeistapausten lkm.}}$$

- Tapahtuman A **geometrinen todennäköisyys**

$$P(A) = \frac{\text{suotuisan osuuden suuruus}}{\text{koko geom. kuvion suuruus}}$$

21. Koulussa on 105 oppilasta. Sählyä harrastaa 40 oppilasta. Tyttöjä on 52, joista 12 harrastaa sählyä. Millä todennäköisyydellä oppilas on
    a) tyttö, joka ei harrasta sählyä
    b) poika, joka harrastaa sählyä
    c) poika, joka ei harrasta sählyä?

22. Ravintolan menu muodostuu seuraavista vaihtoehdoista:
    alkuruuat: keitto, salaatti
    pääruuat: pihvi, uunikala
    jälkiruuat: torttu, sorbetti
    Krista valitsee umpimähkään yhden jokaisesta ryhmästä. Millä todennäköisyydellä hän syö keiton, uunikalan ja sorbetin?

23. Kilpailun viimeistä edelliseen vaiheeseen on selvinnyt viisi henkilöä, mukana mm. Nea ja Leevi. Tästä joukosta arvotaan kaksi henkilöä loppukilpailuun. Millä todennäköisyydellä
    a) Nea pääsee loppukilpailuun
    b) Leevi pääsee loppukilpailuun, mutta Nea ei?

24. Millä todennäköisyydellä satunnaisesti valittu henkilö on syntynyt
    a) maanantaina
    b) tammikuussa
    c) klo 12–16 välisenä aikana?

25. 300 hehkulampun palamisajat testattiin ja saatiin seuraavat tulokset:

Palamisaika (h)	Lamppuja
0–300	12
301–600	145
601–900	120
901–	loput

    Millä todennäköisyydellä satunnaisesti valittu lamppu, joka on jo palanut
    a) 300 tuntia, jaksaa vielä palaa 300 tuntia
    b) 600 tuntia, toimii yli 900 tuntia?

26. Suomen pinta-ala on noin 337 000 km^2. Millä todennäköisyydellä maahan putoava meteoriitti putoaa Suomeen, kun maapallon säde on noin 6370 km?

27. Noppaa heitetään kaksi kertaa. Mikä on todennäköisyys, että jälkimmäisellä heitolla saadaan suurempi silmäluku kuin edellisellä? Jos noppaa heitetään kolme kertaa ja ensimmäisellä heitolla on saatu kolmonen, mikä on todennäköisyys, että toisella heitolla saadaan suurempi silmäluku kuin ensimmäisellä ja kolmannella heitolla suurempi kuin toisella? [K2002, 9]

2.2 Komplementti ja kertolaskusääntö

- Tapahtuman A **vastatapahtumaa** $\overline{A}$ kutsutaan **komplementtitapahtumaksi**.
- **Todennäköisyys vastatapahtuman avulla**:

$$P(A) = 1 - P(\overline{A})$$

- Satunnaisilmiön tapahtumat A ja B ovat **riippumattomia**, jos tapahtumien todennäköisyyksiin ei vaikuta se, tapahtuuko myös toinen vai ei.
- **Riippumattomien tapahtumien todennäköisyys**:

$$P(A \text{ ja } B) = P(A) \cdot P(B)$$

28. Malarian vaarallisin muoto aiheutuu *Plasmdium Falciparum* -alkueläimestä. Ehkäisevästä hoidosta huolimatta 30 % tartunnalle altistuneista sairastuu tähän malariaan. Turistiryhmä, jossa on kahdeksan henkilöä, matkustaa malaria-alueelle, jossa kaikki altistuvat taudille. Millä todennäköisyydellä
a) kukaan ei sairastu
b) ainakin yksi saa malarian?

29. Lilli ostaa 6 arpaa. Arvoista joka neljäs arpa voittaa. Millä todennäköisyydellä Lilli voittaa
a) ainakin yhdellä arvalla
b) korkeintaan viidellä arvalla?

30. Heitetään noppaa viisi kertaa. Laske todennäköisyys, että jokaisella heitolla saadaan
a) kuutonen
b) korkeintaan neljä.

31. Geenivirheen aiheuttama sirppisoluanemia on Afrikassa melko yleistä. Erään maakunnan ihmisistä joka kolmas sairastaa anemiaa. Millä todennäköisyydellä tältä alueelta satunnaisesti valitun kahdeksan henkilön joukossa on ainakin yksi tautia sairastava?

32. Eräässä laskinmallissa kahdessa prosentissa laskimista on tyyppivirhe. Matematiikan kurssin opiskelijoista 26:lla on tämä laskin. Millä todennäköisyydellä yksikään laskimista ei ole viallinen?

33. Korttipakasta vedetään peräkkäin neljä korttia, joita ei palauteta pakkaan. Millä todennäköisyydellä
a) saadaan vain punaisia maita
b) vain kuvakortteja
c) kaikki kortit ovat eri maata?

34. Villellä on taskussaan 12 yhden euron kolikkoa. Osa kolikoista on suomalaisia, osa saksalaisia. Ville nostaa taskustaan umpimähkään kaksi kolikkoa. Molemmat ovat suomalaisia. Tämän tapahtuman todennäköisyys on $\frac{5}{33}$. Kuinka monta saksalaista kolikkoa taskussa oli?

35. Tietokonepelissä heitetään virheellistä noppaa, jossa nopan silmälukujen todennäköisyydet ovat suoraan verrannollisia silmälukuihin. Mitkä ovat eri silmälukujen todennäköisyydet? Mikä on todennäköisyys saada kahdella heitolla kaksi kuutosta? [S2001, 9]

2.3 Yhteenlaskusääntö

- Satunnaisilmiöön liittyvät tapahtumat A ja B ovat erillisiä, **jos ne ovat toisensa poissulkevia**.

- **Erillisten tapahtumien yhteenlaskusääntö**:
  $P(A \text{ tai } B) = P(A) + P(B)$

- **Tapahtumien yleinen yhteenlaskusääntö**:
  $P(A \text{ tai } B) = P(A) + P(B) - P(A \text{ ja } B)$

- **Tapahtuu ”*A* tai *B*” tarkoittaa**, että
  – A tapahtuu tai
  – B tapahtuu tai
  – A ja B (molemmat) tapahtuvat.

36. Laatikossa on neljä kultaista, viisi hopeista ja kolme pronssista korvakorua. Kun kaksi korua nostetaan umpimähkään, millä todennäköisyydellä saadaan
    a) hopeiset korut
    b) samasta metallista valmistetut korut?

37. Koulun 358 opiskelijasta tyttöjä on 280, joista 85 on ruotsinkielisiä. Pojista ruotsinkielisiä on 52. Millä todennäköisyydellä
    a) satunnaisesti valittu opiskelija on suomenkielinen tai poika
    b) kahdesta satunnaisesti valitusta opiskelijasta ainakin toinen on ruotsinkielinen?

38. Black Jack -peliä pelataan 52 kortin pakalla. Kaikkien kuvakorttien pistearvo on 10, ässät voivat olla pistearvoltaan 1 tai 11 ja muiden korttien pistearvot ovat normaalit. Pelin tavoitteena on saada pakasta nostettujen korttien pistearvoksi 21.
    a) Mikä on todennäköisyys saada ”Black Jack” eli kahdella kortilla pistearvo 21?
    b) Pelaaja nostaa ensimmäisellä kortilla kuvan ja toisella kortilla kolmosen. Millä todennäköisyydellä kolmannen kortin jälkeen pistearvo ylittää 21?

2.4 Tuloperiaate ja kombinaatio

- **Tuloperiaate**: Jos kokonaisuus muodostuu usean peräkkäisen vaiheen kautta ja jokaisessa vaiheessa voidaan suorittaa tietty määrä valintoja, on erilaisten kokonaisuuksien määrä laskettavissa kertolaskun avulla.

- **Kertoma**: n alkiota voidaan asettaa jonoon $n!$ erilaisella tavalla.

- **Kombinaatio**: n alkion joukosta voidaan valita k kappaletta $\binom{n}{k}$ erilaisella tavalla.

39. Tanssikurssille osallistuu kuusi poikaa ja kuusi tyttöä.
    a) Kuinka monta erilaista tyttö-poika-paria osallistujista voidaan muodostaa?
    b) Osallistujat asetetaan jonoon siten, että joka toinen on tyttö, joka toinen on poika. Kuinka monta erilaista jonoa on mahdollista muodostaa, kun jonon ensimmäinen on poika?

40. Bridge on joukkuepeli, jossa on neljä pelaajaa. Bridgekerhon peli-iltaan saapuu 12 henkilöä.
    a) Kuinka monta erilaista peliseuruetta voidaan muodostaa?
    b) Myös Tupu, Hupu ja Lupu ovat saapuneet paikalle. Jos peliseurueet arvotaan, millä todennäköisyydellä he pelaavat samassa pöydässä?

41. Sanan YLIOPPILAS kymmenestä kirjaimesta otetaan umpimähkään kolme. Millä todennäköisyydellä
    a) ensiksi otettu kirjain on vokaali
    b) otetuista kirjaimista vain yksi on vokaali
    c) otetuista kolmesta kirjaimesta voidaan muodostaa sana ILO? [K2001, 10]

42. Viking-lotossa arvotaan kuusi numeroa ja kaksi lisänumeroa 48 numerosta. Pelaaja tekee yhden kuuden numeron lottorivin. Millä todennäköisyydellä hän saa
    a) neljä oikein
    b) ainakin viisi oikein?

2.5 Binomitodennäköisyys

- Tapahtumaa, jossa samaa koetta toistetaan useita kertoja, kutsutaan **toistokokeeksi**.

- Toistokokeen todennäköisyys eli **binomitodennäköisyys:** $\binom{n}{k} p^k q^{n-k}$
  - p on kokeen onnistumisen todennäköisyys.
  - $q = 1 - p$ on kokeen epäonnistumisen todennäköisyys.

43. Oletetaan, että eräässä kunnassa syntyy ensi vuonna 50 lasta. Lasten sukupuolet ovat toisistaan riippumattomia. Todennäköisyys sille, että syntyvä lapsi on poika, on 0,513. Millä todennäköisyydellä syntyy täsmälleen yhtä monta tyttöä ja poikaa?

44. Heitetään noppaa kuusi kertaa. Kannattaako lyödä vetoa sen puolesta, että kuutonen saadaan ainakin kaksi kertaa?

45. Suomalaisista miehistä 8 % on punavihervärisokeita. Millä todennäköisyydellä umpimähkään valitussa kahdeksan mieshenkilön ryhmässä on
    a) tasan kolme punavihervärisokeaa
    b) korkeintaan kaksi punavihervärisokeaa
    c) vähintään seitsemän punavihervärisokeaa?

1. Harjoituskoe

1. Jokainen opiskelija heitti matematiikan tunnilla noppaa kerran. Heittotulokset on koottu seuraavaan taulukkoon.
   a) Kuinka monta opiskelijaa matematiikan tunnilla on?
   b) Määritä silmäluvun moodi, mediaani ja keskiarvo.
   c) Laske silmäluvun keskihajonta.

Silmäluku	Heittojen lukumäärä
1	9
2	7
3	5
4	4
5	8
6	6

2. Kolme tyttöä heittää 6-tahkoista noppaa, ja kolme poikaa heittää puolestaan 12-tahkoista noppaa. Millä todennäköisyydellä
   a) tytöt saavat jokainen omalla heitollaan saman silmäluvun
   b) pojat saavat jokainen omalla heitollaan vähintään silmäluvun 6?

3. Saara heittää kolikkoa 10 kertaa. Millä todennäköisyydellä hän saa kuusi kruunaa?

4. Aaro jakoi korttipakan 52 korttia maittain neljään erilliseen pinoon.
   a) Aaro ottaa jokaisesta pinosta yhden kortin. Montako erilaista neljän kortin ryhmää hän voi muodostaa?
   b) Kaikki kortit yhdistetään kokonaiseksi pakaksi ja pakka sekoitetaan. Aaro ottaa kokonaisesta pakasta kahdeksan korttia palauttamatta niitä pakkaan. Millä todennäköisyydellä korteissa on vähintään yksi ässä?

5. Taulukkoon on koottu Suomessa vuonna 2012 olevat kesämökit, joiden pinta-ala on 20–99 m^2.
   a) Muodosta kaikki frekvenssijakaumat.
   b) Kuinka monta prosenttia kesämökeistä on alle 60 m^2?
   c) Määritä aineiston mediaaniluokka.
   d) Havainnollista suhteellista frekvenssiä sopivalla kuvaajalla.

Pinta-ala (m^2)	*f*
20–39	133 823
40–59	61 571
60–79	27 967
80–99	27 105

6. Lauri tarkasteli korttipakkojen massoja. Hän havaitsi, että korttipakkojen massa oli keskimäärin 115 g ja keskihajonta 5,3 g. Jos massojen oletetaan jakaantuvan likimain normaalisti, millä todennäköisyydellä Laurin kouluun tuoman korttipakan massa on
   a) yli 118 g
   b) yli 110 g, mutta alle 120 g?

2. Harjoituskoe

1. Maastokisaan osallistui 74 henkilöä. Kisassa jaettiin pisteitä seuraavasti:

Pisteitä	2	3	4	5	6	7	8	9	10
Osallistujia	4	3	7	17	10	9	17	5	2

   a) Kuinka monta prosenttia kisaan osallistuneista sai vähintään 7 pistettä?
   b) Määritä pistemäärien keskiarvo, mediaani ja moodi.
   c) Mikä oli pistemäärien keskihajonta?
   d) Poikkeaako 10 pisteen tulos merkitsevästi keskiarvosta?

2. Erään paikkakunnan kesäkuun lämpötilat (°C) on koottu taulukkoon:

10,6	15,3	19,1	20,0	22,0	21,5	26,4	27,6	27,3	20,2
27,5	25,0	10,7	22,8	22,5	15,9	23,7	22,7	14,3	10,9
15,7	22,9	25,5	20,9	19,7	24,1	11,2	11,3	12,9	23,2

   Luokittele aineisto tasavälisiin luokkiin siten, että ensimmäisenä luokkana on 10–13 (°C) ja muodosta kaikki frekvenssijakaumat. Havainnollista luokittelemaasi aineistoa suhteellisen summafrekvenssin kuvaajan avulla.

3. Siemensäkissä on 40 % siemeniä R, joiden itämistodennäköisyys on 0,9. Loput ovat siemeniä S, joiden itämistodennäköisyys on 0,8. Säkistä nostetaan umpimähkään kaksi siementä. Millä todennäköisyydellä
   a) molemmat nostetut siemenet itävät
   b) ainakin toinen siemenistä itää?

4. Ammattioppilaitos ottaa 120 uutta opiskelijaa. Pekka jää neljännelle varasijalle. Millä todennäköisyydellä Pekka saa kuitenkin opiskelupaikan, kun keskimäärin 7 % peruuttaa opiskelupaikkansa?

5. Erään koulun oppilaiden pituus on likimain normaalisti jakautunut. Pituuksien keskiarvo on 160 cm ja keskihajonta 20 cm. Kuinka monta prosenttia oppilaista on
   a) lyhyempiä kuin 145 cm
   b) pituudeltaan 170–180 cm?

6. Kuningas Arthurin ritarijoukko on vuosien varrella huvennut. Jäljellä on viisi pyöreän pöydän ritaria ja Arthur itse. Koko joukko järjestäytyy kuuluisan pyöreän pöydän ääreen.
   a) Kuinka moneen erilaiseen järjestykseen tämä kuuden hengen joukko voi asettua, kun Arthur itse istuu aina samalla paikalla eli valtaistuimellaan?
   b) Arthur haluaa lähettää kaksi ritareistaan naapurivaltion kuninkaan järjestämiin turnajaisiin. Kuinka monta erilaista kahden hengen joukkuetta voidaan muodostaa?
   c) Millä todennäköisyydellä sir Lancelot pääsee turnajaisiin, kun lähtijät arvotaan?

3. Harjoituskoe

1. Taulukossa on erään luokan oppilaiden pituudet luokiteltuina. Määritä pituuden keskiarvo, moodi, mediaani ja keskihajonta. Kuvaa frekvenssijakaumaa graafisesti sopivalla tavalla.

Pituus (cm)	*f*
155–159	7
160–164	6
165–169	6
170–174	8
175–179	1
180–184	2

2. Täydennyskoulutuksessa on 50 opiskelijaa, joista 33 käyttää sähköpostin lukemiseen puhelinta ja 15 puhelimen lisäksi myös iPadia. Opiskelijoista seitsemän on sellaisia, jotka eivät käytä sähköpostin lukemiseen kumpaakaan mainituista laitteista. Millä todennäköisyydellä opiskelija lukee sähköpostinsa
   a) pelkästään iPadilla
   b) puhelimella tai iPadilla
   c) pelkästään puhelimella?

3. Matematiikan ryhmässä on 26 oppilasta. Jokaisen oppilaan laukussa on keskimäärin viisi kirjaa.
   a) Millä todennäköisyydellä oppilas umpimähkään, laukkuunsa katsomatta, tarttuu heti oikeaan kirjaan?
   b) Millä todennäköisyydellä ryhmän kaikki oppilaat valitsevat heti oikean kirjan?
   c) Luokkaan tullessaan opettaja vetää tavallisesta korttipakasta (52 korttia) yhden kortin. Jos kortti on ässä tai hertta 10, hän kuulustelee läksyn kirjallisesti. Laske tapahtuman todennäköisyys.

4. Erään tutkimuksen mukaan 60 % ihmisistä käyttää jonkinlaisia silmälaseja ja vain 40 % selviää ilman laseja kaikissa tilanteissa. Millä todennäköisyydellä satunnaisesti valitussa 8 henkilön ryhmässä on
   a) viisi silmälaseja käyttävää henkilöä
   b) ainakin kaksi silmälaseja käyttävää henkilöä?

5. Tieosuudella, jolla nopeusrajoitus oli 80 km/h, oli mittausten mukaan ajoneuvojen keskinopeus 77,0 km/h ja keskihajonta 7,0 km/h. Kuinka monta prosenttia autoilijoista olisi voinut saada rikesakon (nopeus yli 80 km/h, mutta alle 95 km/h), kun nopeuden jakauma oli normaali?

6. Kahviautomaatista saatavan annoksen tilavuus on normaalisti jakautunut siten, että annoksien keskiarvo on 200 cm^3 ja keskihajonta 4 cm^3. Kuinka suuri on kahvimukin oltava, jotta ylivalumisen todennäköisyys on alle 10 %?

Projektityö: Projektityön ohjeita

Projektityönä tehtävä tilastollinen tutkimus tulee suunnitella kunnolla. Tutkimuksen aiheen ja tutkittavien muuttujien valintaan kannattaa käyttää aikaa samoin kuin tiedon keräämistavan valintaan. Näitä pohdittaessa on syytä muistaa, että kun tutkittavat muuttujat on päätetty ja havaintoaineisto kerätty, ei havaintoaineistoa voida enää muuttaa.

Tutkimuksen suunnittelu

Suunnitteluvaiheessa mieti vastaukset seuraaviin kysymyksiin:

- Mitä tutkitaan eli mikä on tutkimuksen aihe?
- Mitkä ovat tutkittavat muuttujat?
  - Valitse ainakin neljä muuttujaa.
  - Valitse erilaisia muuttujia (diskreettejä/jatkuvia).
  - Valitse ainakin kaksi jatkuvaa muuttujaa.
- Miten aineisto kerätään?
  - Jos teet kyselylomakkeen, mieti kaikki tarvittavat kysymykset huolella.
- Mikä on tutkimuksen aikataulu ja työnjako?
  - Tunnuslukujen laskeminen symbolisella laskimella tai taulukkolaskentaohjelmalla nopeuttaa tutkimusta.
  - Jos projekti tehdään parityönä, kannattaa kirjata ylös, mitä kumpikin työparista tekee. Näin työt on helpompi jakaa tasaisesti.

Tietojen analysointi

Analysointia varten havaintoaineiston muuttujille muodostetaan frekvenssijakaumia ja lasketaan tilastollisia tunnuslukuja. Ennen jakaumien tekemistä jatkuvat muuttujat on luokiteltava.

Jotta projektityö etenisi samassa järjestyksessä kuin kurssilla käsiteltävät asiat, kannattaa havaintoaineiston käsittely ja tietojen analysointi tehdä seuraavassa järjestyksessä:

1. **Frekvenssijakaumat**
   - Luokittele tarvittaessa muuttujat (jatkuvat muuttujat).
   - Muodosta frekvenssijakaumat (f, f %, sf, sf %).
   - Piirrä jakaumien kuvaajat.
   - Analysoi aineistoa frekvenssijakaumien avulla.

2. **Tunnusluvut**
   - Määritä muuttujille moodi, mediaani ja keskiarvo.
   - Analysoi aineistoa keskilukujen avulla.
   - Määritä muuttujille hajontaluvut (keskihajonta, vaihteluväli ja sen pituus).
   - Analysoi aineistoa hajontalukujen avulla.
   - Ekstraosan avulla voit lisäksi tutkia kahden jatkuvan muuttujan riippuvuutta yhteisjakauman ja korrelaatiokertoimen avulla.

Huom!
Kaikille muuttujille ei välttämättä voi laskea kaikkia tunnuslukuja.

Raportin kirjoittaminen

Projektityöstä kootaan lopuksi raportti. Raportti voi muodostua seuraavista luvuista:

1. **Johdanto**
   - Mitä asiaa tutkitaan?
   - Mihin kysymyksiin halutaan vastaus?
   - Miten aineisto on kerätty (keräämistapa)?
   - Mitkä ovat tutkimuksen havaintoyksiköt?
   - Mitkä ovat tutkimuksen muuttujat?
   - Ovatko muuttujat jatkuvia vai diskreettejä?

2. **Frekvenssijakaumat**
   - Liitä muuttujien frekvenssijakaumat ja jakaumien kuvaajat.
   - Analysoi tilannetta frekvenssijakaumien perusteella.

3. **Tunnusluvut**
   - Määritä muuttujille keskiluvut (moodi, mediaani, keskiarvo).
   - Analysoi tilannetta keskilukujen avulla.
   - Laske muuttujille hajontaluvut (keskihajonta, vaihteluväli).
   - Analysoi tilannetta hajontalukujen avulla.

4. **Yhteisjakauma**
   - Piirrä kahden vähintään välimatka-asteikollisen muuttujan yhteisjakauman kuvaaja.
   - Laske korrelaatiokerroin.
   - Sovita regressiokäyrä.
   - Analysoi tilannetta yhteisjakauman ja korrelaatiokertoimen avulla.

Yhteisjakauman tarkastelun voi jättää käsittelemättä.

Lisätietoa muuttujien asteikoista ja yhteisjakaumasta Ekstra-osassa.

5. **Yhteenveto**
   - Kokoa yhteen tärkeimmät tulokset.
   - Mitä johtopäätöksiä voit tutkimuksesi perusteella tehdä?

6. **Liitteet**
   - Kysymys/haastattelulomake
   - Erilliset kuvaajat

Havaintoaineiston kerääminen

Havaintoaineisto voidaan kerätä monella eri tavalla. Tapa valitaan sen mukaan, millaista tietoa halutaan kerätä.

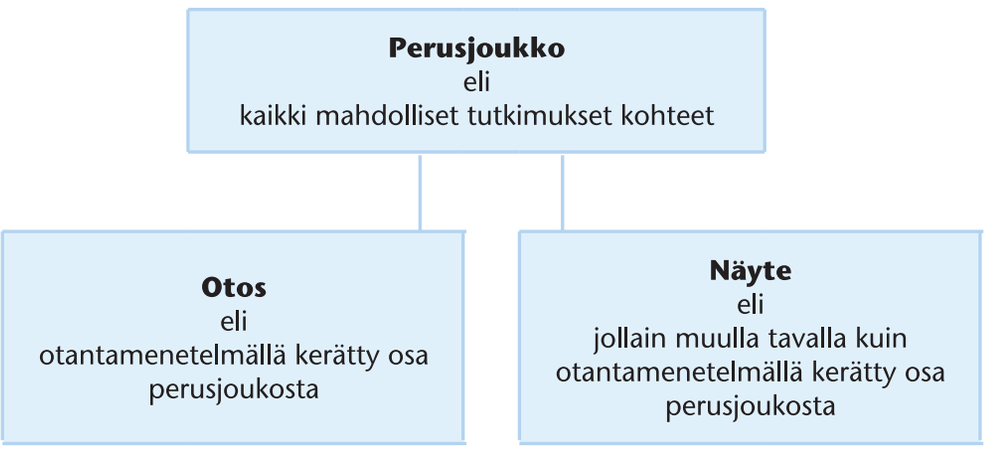
Keräämistapa	Miten tieto kerätään?	Esimerkki tutkimuksesta
Kysely	Havaintoaineistoon valittu henkilö täyttää kyselylomakkeen itse.	kauppojen ja ravintoloiden asiakastyytyväisyys
Haastattelu	Haastattelija täyttää lomakkeen haastateltavan vastausten perusteella.	kadulla tehtävät mielipidekyselyt
Havainnointi	Havainnoija kerää tietoja omien havaintojensa (ei haastattelun) perusteella.	käyttäytymistutkimukset
Systemaattinen koejärjestely	Tietoa kerätään kontrolloiduissa olosuhteissa suoritetun kokeen perusteella.	Makutesti: mikä seuraavista kolmesta virvoitusjuomasta maistuu parhaimmalta?
Tietokanta	Tieto on jo kerätty.	koulun oppilastietojen perusteella suoritettava tutkimus

Kysymysten laadinta on eräs tutkimuksen tärkeimmistä vaiheista. Huonoilla kysymyksillä ei saa tarvittavaa tietoa selville ja koko tutkimus voi mennä pieleen. Esitettävien kysymysten tulisi olla selkeitä ja yksiselitteisiä.

Tilastollisissa tutkimuksissa havaintoyksiköt kerätään yleensä jollain otantamenetelmällä. Seuraavaan taulukkoon on koottu muutamia otantamenetelmiä. Taulukon esimerkkisarakkeessa tutkimuksen kohteena ovat Suomen lukiolaiset.

Otantamenetelmä	Miten otanta suoritetaan?	Esimerkki
Yksinkertainen satunnaisotanta	• Havaintoyksiköt poimitaan perusjoukosta arpomalla. • Jokaisella havaintoyksiköllä on yhtä suuri mahdollisuus tulla valituksi.	• Otokseen poimitaan haluttu määrä lukiolaisia satunnaisesti.
Systemaattinen otanta	• Perusjoukosta poimitaan havaintoyksiköitä tietyin välein. • Aloituskohta voidaan arpoa.	• Otokseen otetaan mukaan esimerkiksi joka 20. lukiolainen.
Ositettu otanta	• Perusjoukko on jakaantunut erilaisiin luokkiin jonkin yleisen ominaisuuden, kuten iän tai sukupuolen, perusteella. • Otokseen poimitaan satunnaisotannalla jokaisesta luokasta tietty määrä havaintoyksiköitä.	• Otokseen valitaan mukaan haluttu määrä sekä lukiossa opiskelevista tytöistä että pojista.
Ryväsotanta	• Perusjoukko on jakaantunut ryhmiin eli rypäisiin jonkin paikallisen ominaisuuden, kuten asuinpaikan tai työpaikan, perusteella. • Tietty määrä kokonaisia ryhmiä poimitaan satunnaisotannalla.	• Kymmenen lukiota valitaan satunnaisesti mukaan otokseen.

Jos havaintoyksiköt poimitaan perusjoukosta jollain muulla tavalla kuin otantamenetelmällä, kerättyä aineistoa kutsutaan **näytteeksi** eli **harkintaotokseksi**. Otoksen avulla saadut tulokset voidaan useimmiten yleistää koskemaan koko perusjoukkoa. Näytteeseen perustuvien tulosten yleistäminen on ongelmallisempaa, koska näyte saattaa olla hyvin valikoitunut ja huonosti perusjoukkoa edustava.



Tutkimuksen kohde Suomen lukiolaisten koulumatkan pituus	
Kokonaisaineisto	Havaintoaineistona ovat kaikki Suomen lukiolaiset.
Otos	Havaintoaineistona ovat satunnaisesti valitut lukiolaiset.
Näyte	Havaintoaineistona ovat tietyn koulun lukiolaiset.

Tehtäviä

1. Tilastotieteilijä haluaa selvittää, miten Suomen lukiolaiset käyttävät rahaa vapaa-aikanaan. Kerro esimerkki siitä, miten havaintoaineisto kannattaa koota, jos havaintoaineisto on
   a) otos b) näyte c) populaatio.

2. Mitkä seuraavista ovat otantatutkimuksia ja mitkä kokonaistutkimuksia?
   a) matkapuhelintehtaan laadunvalvonta
   b) ravintolan tekemä asiakastyytyväisyystutkimus
   c) kansanedustajien mielipidetutkimus
   d) eduskuntavaalien äänestystuloksen ennustus

3. Onko tutkimus otantatutkimus vai kokonaistutkimus?
   a) lentoyhtiön asiakastyytyväisyyskysely kotimaan aamulennoilla
   b) lomamatkan järjestäjän asiakastyytyväisyyskysely
   c) lentokentälle saapuvien matkustajien asiakastyytyväisyyskysely

4. Millä tavalla keräisit tietoja sellaista tutkimusta varten, joka selvittää
   a) kuinka suurta opintotukea lukiolaiset saavat
   b) mitä virvoitusjuomia lukioikäiset juovat
   c) ovatko opiskelijat tyytyväisiä koulun koeviikkokäytäntöön
   d) kuinka moni etuilee ruokalan jonossa?
   Perustele vastauksesi.

5. Millä tavalla tieto kannattaa kerätä seuraavissa tutkimuksissa?
   a) kaupan asiakkaiden asiakastyytyväisyystutkimus
   b) kaupan työntekijöiden palkkatasotutkimus
   c) kahden tuotteen makuvertailu
   d) kaupan asiakkaiden kassanvalintatutkimus
   Perustele vastauksesi.

6. Millä tavalla keräisit tietoa seuraavissa tutkimuksissa?
   a) opiskelijoiden myöhästyminen oppitunnilta
   b) opettajien ja oppilaiden mielipide kouluruoasta
   c) pulpetin valitseminen tyhjässä luokassa
   Perustele vastauksesi.

Muuttujien luokittelu

Havaintoaineiston muuttujat voidaan luokitella kuvaustavan tai mitta-asteikon mukaan. Muuttujan kuvaustapoja on kahdenlaisia:

– Jos muuttujan luonnollinen kuvaustapa on reaaliluku, muuttujaa sanotaan **kvantitatiiviseksi muuttujaksi**.
– Jos taas muuttujaa ei kuvata reaaliluvulla tai luku on vain tietty koodiarvo, muuttujaa sanotaan **kvalitatiiviseksi muuttujaksi**.

ESIMERKKI 1

Mitkä seuraavan kyselylomakkeen muuttujat ovat kvalitatiivisia ja mitkä kvantitatiivisia?

Sukupuoli ☐ mies ☐ nainen

Ikä _____

Pituus _____

Mielipide kouluruoasta
1 Ruoka on huonoa.
2 Ruoka on ihan hyvää.
3 Ruoka on erinomaista.

RATKAISU

Sukupuoli on kvalitatiivinen muuttuja, koska sen luonnollinen kuvaustapa ei ole luku.

Ikä ja pituus ovat kvantitatiivisia muuttujia, koska niiden luonnollinen kuvaustapa on luku.

Mielipide kouluruoasta on kvalitatiivinen muuttuja, koska sen luonnollinen kuvaustapa on laatusana.

VASTAUS

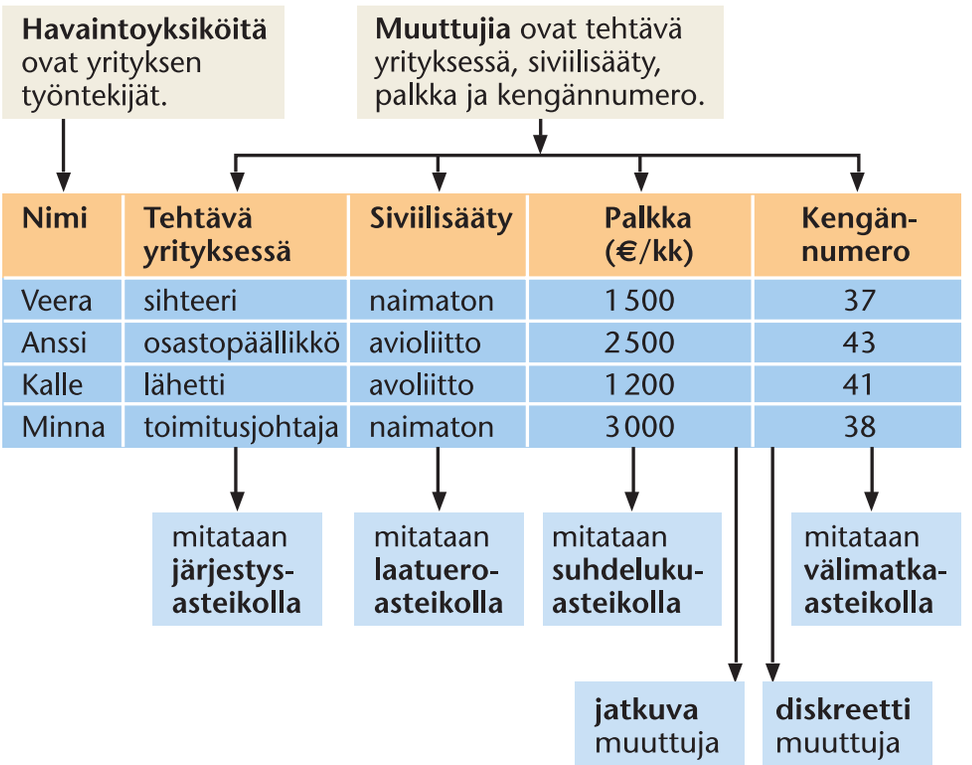
Sukupuoli ja mielipide kouluruoasta ovat kvalitatiivisia muuttujia, ikä ja pituus kvantitatiivisia muuttujia.

Mitta-asteikon mukaan muuttujat voidaan jaotella neljään luokkaan, jotka on koottu seuraavaan taulukkoon. Mitta-asteikot rakentuvat hierarkkisesti, joten taulukkoon on koottu vain ne muuttujan ominaisuudet, jotka poikkeavat edellisistä asteikoista. Muuttujan mittaustavalla on merkitystä tunnuslukuja laskettaessa.

Mitta-asteikko	Millainen muuttuja?	Esimerkki muuttujasta
Laatuero- eli nominaaliasteikko	• Muuttuja kuvaa havaintoyksiköiden välistä erilaisuutta tai samanlaisuutta. • Muuttujan arvot voidaan luokitella, mutta niitä ei voida asettaa esimerkiksi suuruusjärjestykseen.	• sukupuoli • siviilisääty
Järjestys- eli ordinaaliasteikko	• Muuttujan arvot voidaan järjestää suuruusjärjestykseen. • Laskutoimitukset eivät ole mielekkäitä.	• matematiikan osaaminen asteikolla 1. en osaa … 5. osaan erinomaisesti
Välimatka- eli intervalliasteikko	• Nollakohta voidaan sopia erikseen. • Useimmat laskutoimitukset ovat mahdollisia. • Muuttujan arvojen välistä suhdetta ei voida määrittää yksikäsitteisesti (yksiköstä riippumatta).	• lämpötila (yksikkönä esimerkiksi celsius- tai fahrenheitaste) • kalenteriaika
Suhdelukuasteikko	• Absoluuttinen nollakohta on olemassa, eli nollakohtaa ei voida sopia erikseen. • Muuttujan arvojen suhde voidaan määrittää yksikäsitteisesti (yksiköstä riippumatta).	• pituus (yksikkönä esimerkiksi senttimetri tai tuuma) • lämpötila kelvinasteina

Välimatka- ja suhdelukuasteikolliset muuttujat voidaan luokitella myös muuttujan jatkuvuuden mukaan. **Jatkuva muuttuja**, esimerkiksi henkilön pituus, voi saada mitä tahansa arvoja tietyllä välillä. **Epäjatkuva** eli **diskreetti muuttuja**, esimerkiksi henkilön kengännumero, voi saada vain tiettyjä arvoja. Järjestys- ja laatueroasteikolliset muuttujat ovat aina diskreettejä.

Tarkastellaan esimerkkinä erään yrityksen työntekijöistä laadittua tilastoa.



Tehtävä yrityksessä on **järjestysasteikollinen** muuttuja, koska muuttujan arvot voidaan laittaa järjestykseen, mutta niiden väliset laskutoimitukset eivät ole mielekkäitä.

Siviilisääty on **laatueroasteikollinen** muuttuja, koska muuttujien arvoja ei voida järjestää.

Muuttujalla palkka on yksikäsitteinen nollakohta. Lisäksi muuttujan arvojen suhde on aina, yksiköstä riippumatta, sama. Esimerkiksi Minnan palkan suhde Veeran palkkaan on $\frac{3000\,€}{1500\,€} = 2$. Palkka on siis **suhdelukuasteikollinen** muuttuja.

Kengännumero voidaan ilmoittaa usealla eri tavalla. Euroopassa tyypillisimmät tavat perustuvat ranskalaiseen ja englantilaiseen järjestelmään. Suhdetta ei voida määrittää yksikäsitteisesti, joten kengännumero on **välimatka-asteikollinen** muuttuja.

	Minna	Kalle	Suhde
Ranskalainen järjestelmä	38	41	$\frac{38}{41} = 0{,}926\ldots$
Englantilainen järjestelmä	5	7	$\frac{5}{7} = 0{,}714\ldots$

Tehtäviä

7. Onko muuttuja jatkuva vai diskreetti?
   a) auton huippunopeus
   b) auton vaihteiden määrä
   c) auton bensiininkulutus
   d) auton nopeusmittarissa näkyvä suurin nopeus

8. Alla olevaan taulukkoon on kerätty tietoja muutamista EU-maista.
   a) Mitkä aineiston muuttujista ovat kvantitatiivisia ja mitkä kvalitatiivisia?
   b) Mitkä aineiston muuttujista ovat jatkuvia ja mitkä diskreettejä?
   c) Millä mitta-asteikolla muuttujat on mitattu?

	Ala (km^2)	Väkiluku (milj.)	Valtiomuoto	Verojen osuus BKT:sta (%)
Suomi	338 127	5,132	tasavalta	45,9
Ruotsi	449 964	8,858	monarkia	49,3
Saksa	356 805	81,891	liittotasavalta	37,7
Alankomaat	40 844	15,589	monarkia	44,9
Ranska	551 600	58,392	tasavalta	43,1

9. Millä mitta-asteikolla seuraavat muuttujat on mitattu?
   a) henkilön reaktioaika ärsykkeeseen
   b) henkilön yleiskunto asteikolla 1–4 (1 = erinomainen, 4 = huono)
   c) henkilön siviilisääty
   d) henkilön työmatkakustannukset vuodessa

10. Marika teki tutkimusta käytetyistä autoista. Millä mitta-asteikolla hänen havaintoaineistonsa muuttujat on mitattu?

Auton merkki	Ajetut kilometrit	Kuntoarvio	Väri	Rekisteröintivuosi	Hinta-arvio (€)
BMW	34 000	hyvä	harmaa	2011	19 900
Volkswagen	96 000	erinomainen	musta	2009	11 500
Audi	206 000	tyydyttävä	valkoinen	2005	8 000

Frekvenssien kuvaaminen

Frekvenssijakauman muoto riippuu siitä onko tarkasteltava muuttuja diskreetti vai jatkuva. Lisäksi frekvenssijakaumaa ja summafrekvenssijakaumaa kuvataan erilaisilla kuvaajilla.

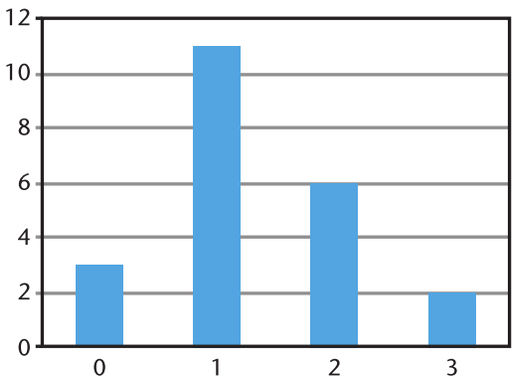
Diskreetti muuttuja

Tarkastellaan ensin diskreetin muuttujan, sisarusten lukumäärä, jakaumien kuvaajia.

Frekvenssijakauma ja suhteellinen frekvenssijakauma

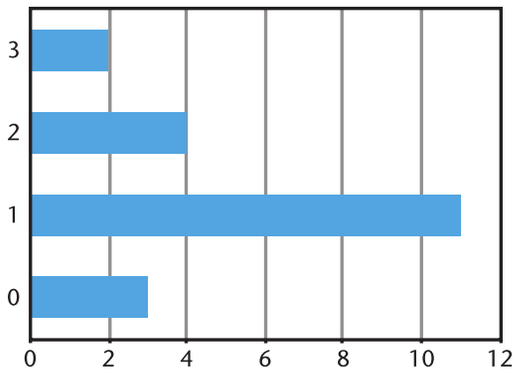
Sisarusten lukumäärä	*f*	*f* %
0	3	14 %
1	11	50 %
2	6	27 %
3	2	9 %

Pylväsdiagrammi



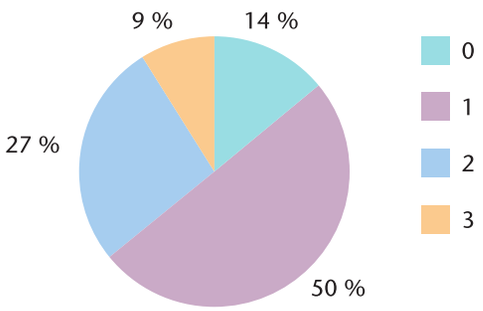
- Pylvään korkeus määräytyy frekvenssin tai suhteellisen frekvenssin mukaan. Tällöin pylväiden on oltava keskenään yhtä leveitä. Jos pylväiden leveys vaihtelee, frekvenssiä kuvataan pylvään pinta-alalla.
- Pylväät piirretään erilleen.

Palkkidiagrammi



- Palkin eli vaakapylvään korkeus määräytyy frekvenssin tai suhteellisen frekvenssin mukaan.
- Palkit piirretään erilleen.

Sektoridiagrammi

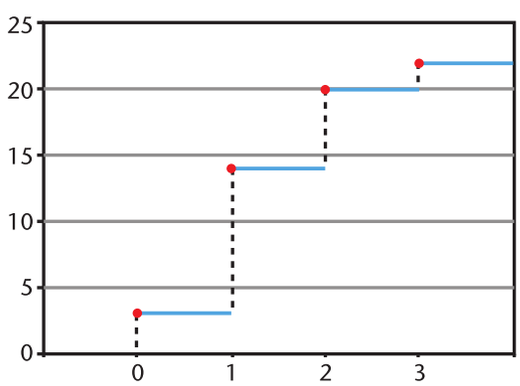


- Sektoridiagrammilla kuvataan suhteellisia frekvenssejä.
- Sektorin keskuskulman suuruus määräytyy suhteellisen frekvenssin mukaan.

Summafrekvenssi- ja suhteellinen summafrekvenssijakauma

	sf
0	3
1	14
2	20
3	22

Porrasdiagrammi



- Summafrekvenssit tai suhteelliset summafrekvenssit merkitään kuvaajaan pisteillä.
- Kertymä pysyy samana aina seuraavaan muuttujan arvoon asti, joten muuttujan arvojen välille piirretään vaakaviiva.
- Vaakaviivat yhdistetään katkoviivoilla.

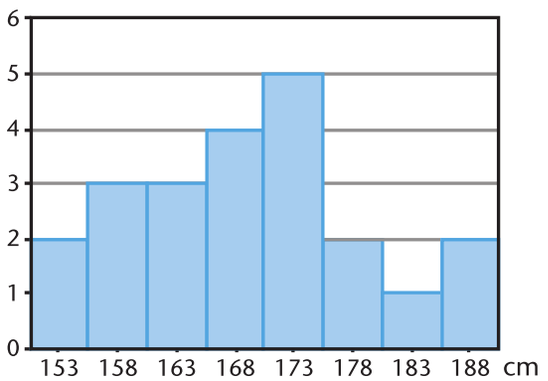
Jatkuva muuttuja

Tarkastellaan jatkuvan muuttujan, pituus, jakaumien kuvaajia.

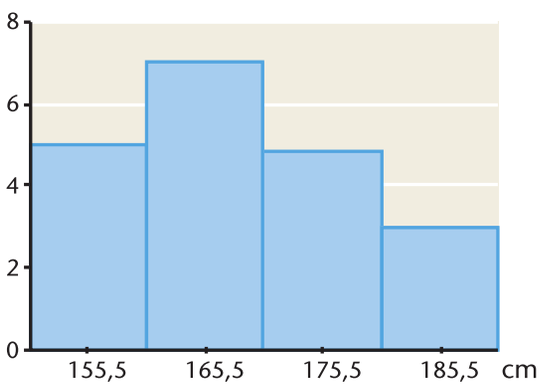
Frekvenssijakauma ja suhteellinen frekvenssijakauma

Histogrammi

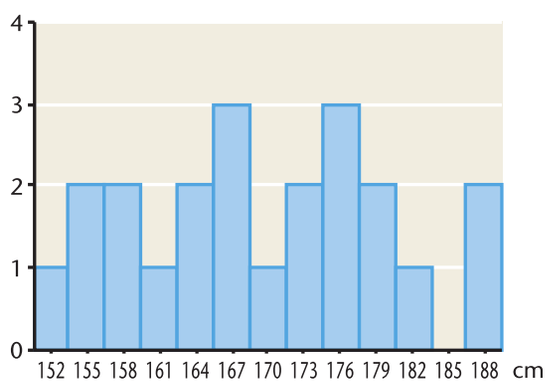
Pituus	151–155	156–160	161–165	166–170	171–175	176–180	181–185	186–190
f	2	3	3	4	5	2	1	2



- Pylvään korkeus määräytyy frekvenssin tai suhteellisen frekvenssin mukaan. Tällöin pylväiden on oltava keskenään yhtä leveitä. Jos pylväiden leveys vaihtelee, frekvenssiä kuvataan pylvään pinta-alalla.
- Pylväät piirretään yhteen.
- Pylvään keskikohdalle merkitään luokkakeskus tai reunoille luokkarajat.

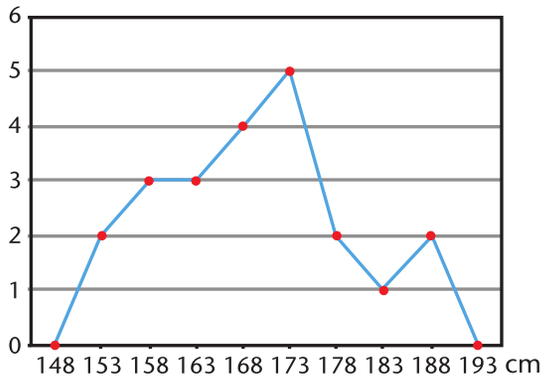


Jatkuvaa muuttujaa kuvattaessa frekvenssijakauman muoto riippuu aineiston luokittelusta. Luokkavälejä suurentamalla luokkien määrä vähenee ja tieto ”tiivistyy”. Samalla yksityiskohdat katoavat.



Luokkavälejä pienentämällä luokkien määrä kasvaa ja yksityiskohdat saadaan paremmin näkyviin, mutta tilanteen kokonaiskuva hämärtyy.

Frekvenssimonikulmio

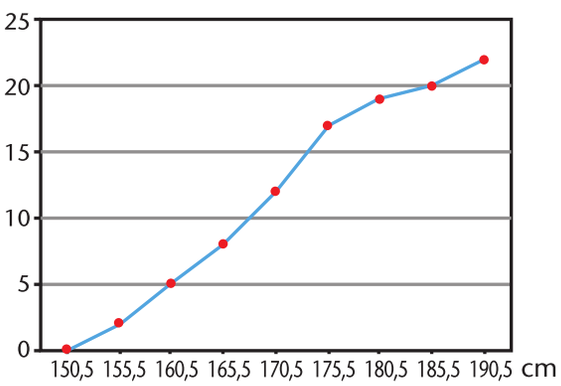


- Luokkakeskusten kohdalle merkitään luokkaa vastaava frekvenssi tai suhteellinen frekvenssi.
- Frekvenssipisteet yhdistämällä saadaan murtoviiva.
- Jakauman hännät päättyvät 0-tasolle.

Summafrekvenssi- ja suhteellinen summafrekvenssijakauma

Summakäyrä eli summapolygoni

Pituus	151–155	156–160	161–165	166–170	171–175	176–180	181–185	186–190
sf	2	5	8	12	17	19	20	22

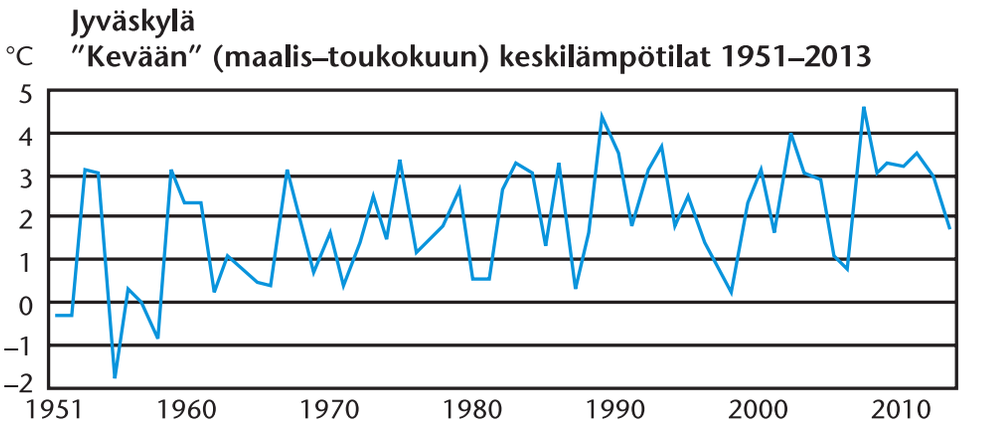


- Koska summafrekvenssit ja suhteelliset summafrekvenssit kertovat tiettyyn arvoon mennessä kertyneiden havaintoyksiköiden määrän, frekvenssien arvot sijoitetaan luokkien todellisten ylärajojen kohdalle.
- Summafrekvenssipisteet yhdistämällä saadaan murtoviiva.
- Jakauma alkaa nollasta.

Aikasarja

Jos samasta tutkimuksen kohteena olevasta asiasta tehdään havaintoja peräkkäisten, yhtä pitkien ajanjaksojen jälkeen, puhutaan **aikasarjasta**. Aikasarjoja kuvattaessa muuttujan arvot sijoitetaan kuvaajaan aikajärjestyksessä. Aikasarjojen kuvaajia käytetään usein erilaisten tulevaisuutta koskevien ennusteiden tekemiseen.

Esimerkiksi Jyväskylän kevään keskilämpötilat vuosina 1951–2013 muodostavat aikasarjan.

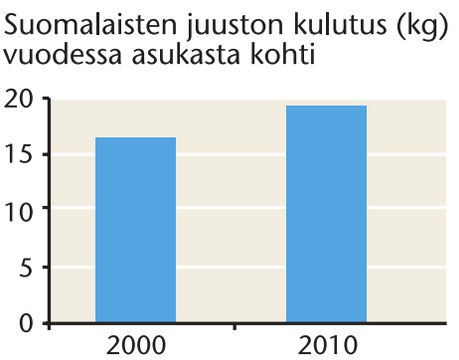


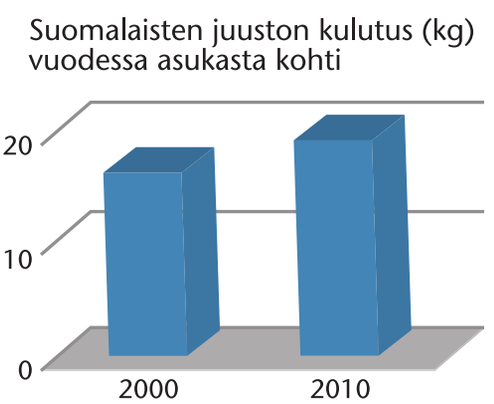
Lähde: Ilmatieteen laitos

Tilastoilla valehtelu

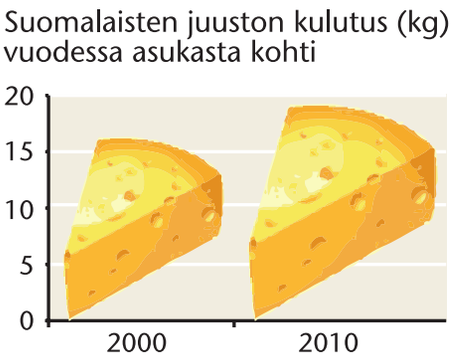
Kun frekvenssijakaumaa havainnollistetaan graafisesti, tarkoitus on tuoda esiin jakauman olennaiset piirteet ja auttaa tekemään oikeita johtopäätöksiä. Kuvaajaa laadittaessa on aina mietittävä tarkasti, mitä sillä halutaan kuvata. Kuvaajilla on helppo valehdella eli esittää asiat niin, että lukijalle syntyy haluttu mielikuva.

Kolmiulotteiset kuvaajat saattavat näyttää hienommilta, mutta useimmiten tiedon luettavuus kärsii.

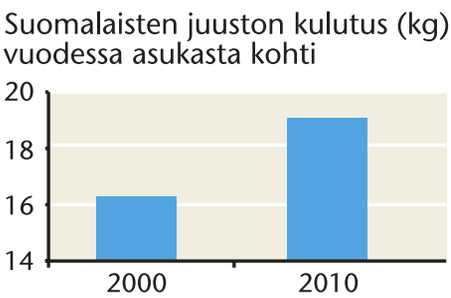




Pylväiden tilalla käytetään usein myös kuvia, jolloin mielikuva todellisista frekvensseistä vääristyy helposti. Jos kuvat tai pylväät eivät ole keskenään yhtä leveitä, frekvenssejä kuvataan kuvan tai pylvään pinta-alalla, ei korkeudella.

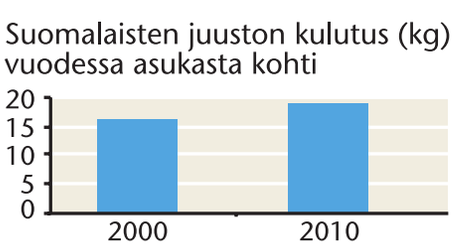


Asteikon katkaiseminen on yksi tapa vaikuttaa kuvaajan muotoon. Kun pylväs- tai viivadiagrammissa pystyakseli katkaistaan eli asteikko alkaakin jostain muusta kuin luvusta nolla, vaihtelut saadaan näyttämään suuremmilta.

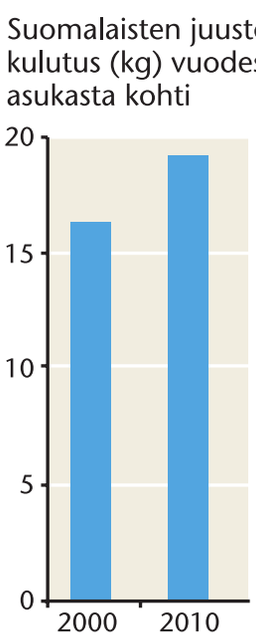


Myös akselien venyttämisellä voidaan vaikuttaa halutun mielikuvan syntymiseen. Asteikkoja venyttämällä tai kaventamalla saadaan muutokset näyttämään joko suuremmilta tai pienemmiltä. Yleensä kuvio on sopusuhtainen silloin, kun akselit ovat yhtä pitkiä tai akselien suhde on sama kuin A4-arkin sivujen suhde.

Vaaka-askelia venyttämällä muutokset näyttävät pienemmiltä.



Pystyakselia venyttämällä muutokset näyttävät suuremmilta.



Tehtäviä

11. Taulukossa on bensiinin (95E) litrahinnat Suomessa vuosina 2005–2012. Piirrä aikasarjan kuvaaja.

Vuosi	2005	2006	2007	2008	2009	2010	2011	2012
Hinta (€/l)	1,22	1,29	1,30	1,41	1,29	1,43	1,56	1,67

Lähde: Tilastokeskus

12. Etsi lehdistä kolme erilaista aikasarjan kuvaajaa.

13. Viereisessä taulukossa on esitetty kahvin kilohinta Suomessa vuosina 2005–2012. Kuvaa hinnan kehitystä trendikäyrällä niin, että hinta näyttäisi muuttuvan
a) mahdollisimman vähän
b) mahdollisimman paljon.

Vuosi	Hinta (€/kg)
2005	2,52
2006	2,58
2007	2,62
2008	2,79
2009	2,78
2010	3,06
2011	4,42
2012	4,11

Lähde: Tilastokeskus

14. Leo oli ottamassa asuntolainaa syksyllä 2013, kun pankkien korot olivat alhaalla. Hän pyysi lainatarjouksia kahdesta eri pankista. Pankit tarjosivat pankin omaan korkoon sidottua lainaa. Pankkien omat korot olivat 2,25 % ja 2,60 %. Piirrä koroista pylväsdiagrammi, jossa esität pankkien korkoprosentit
a) alemman koron b) korkeamman koron
tarjonneen pankin näkökulmasta. Käytä hyväksesi tilastoilla valehtelun keinoja.

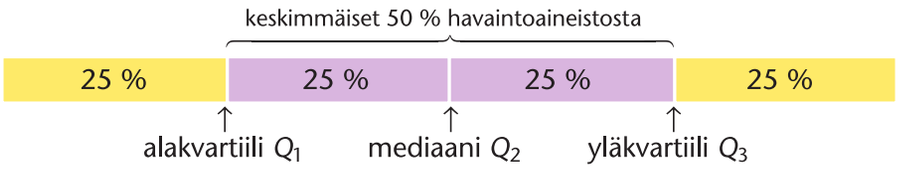
15. Kirjassa on esitetty muutamia tapoja tilastoilla valehteluun. Keksi lisää tapoja.

Kvartiilit

Kvartiilit ovat muuttujan arvojen kertymää kuvaavia sijaintilukuja. Kvartiileja on kolme:

- **alakvartiili** (Q_1)
- mediaani (Md tai Q_2)
- **yläkvartiili** (Q_3).

Kvartiilit jakavat muuttujan arvojen mukaisesti havaintoaineiston neljään yhtä suureen osaan, joista jokaisessa on 25 % havaintoyksiköiden määrästä.



Kvartiilit

Alakvartiili Q_1: muuttujan arvo, jota pienempiä arvoja on 25 %.

Mediaani Q_2: muuttujan arvo, jota pienempiä arvoja on 50 %.

Yläkvartiili Q_3: muuttujan arvo, jota pienempiä arvoja on 75 %.

Hajontaa kuvaamaan voidaan käyttää **kvartiiliväliä** (Q_1, Q_3). Alakvartiilin Q_1 ja yläkvartiilin Q_3 väliin jää keskimmäiset 50 % muuttujan arvoista. Voidaan myös laskea kvartiilivälin pituus, joka on yläkvartiilin ja alakvartiilin erotus $Q_3 - Q_1$.

Seuraavassa taulukossa on erään oppilasryhmän pituuksien frekvenssijakauma.

Luokka	151–155	156–160	161–165	166–170	171–175	176–180	181–185	186–190
f	2	3	3	4	5	2	1	2

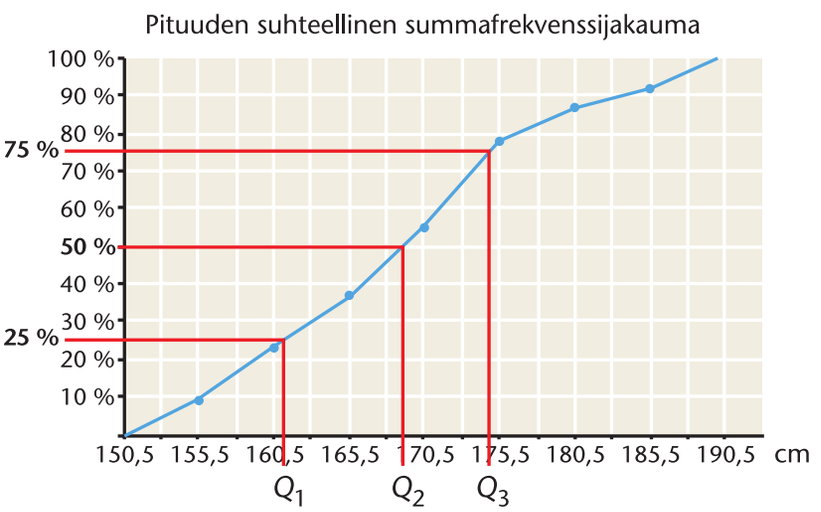
Luokitellusta aineistosta määritetään ensin kvartiililuokat suhteellisen summafrekvenssijakauman avulla.

Luokka	*f*	*f* %	*sf*	*sf* %	
151–155	2	9 %	2	9 %	
156–160	3	14 %	5	23 %	
161–165	3	14 %	8	36 %	Alakvartiililuokka
166–170	4	18 %	12	54 %	
171–175	5	23 %	17	77 %	Yläkvartiililuokka
176–180	2	9 %	19	86 %	
181–185	1	5 %	20	91 %	
186–190	2	9 %	22	100 %	

Alakvartiililuokka on se luokka, jossa suhteellinen summafrekvenssi on ensimmäisen kerran yli 25 %. Tässä tapauksessa alakvartiililuokka on luokka 161 cm – 165 cm. Alakvartiilin likiarvona voidaan käyttää luokan luokkakeskusta eli $Q_1 = 163$ cm.

Yläkvartiililuokka on se luokka, jossa suhteellinen summafrekvenssi on ensimmäisen kerran yli 75 %. Tässä tapauksessa yläkvartiililuokka on luokka 171 cm – 175 cm. Myös yläkvartiilin likiarvona voidaan käyttää luokkakeskusta eli $Q_3 = 173$ cm.

Kvartiileille voidaan määrittää likiarvot graafisesti summakäyrän avulla.



Kuvaajan perusteella alakvartiili on noin 161 cm ja yläkvartiili noin 174 cm. Kvartiiliväli on siis (161 cm, 174 cm) ja kvartiilivälin pituus 174 cm – 161 cm = 13 cm.

Tehtäviä

16. Erään ryhmän poissaoloja tutkittiin kuukauden ajan. Tutkimuksen tulokset on koottu taulukkoon. Mikä on poissaolopäivien
a) alakvartiili b) mediaani c) yläkvartiili?

Poissaolopäivät	*f*
0	8
1	6
2	5
3	7
4	6

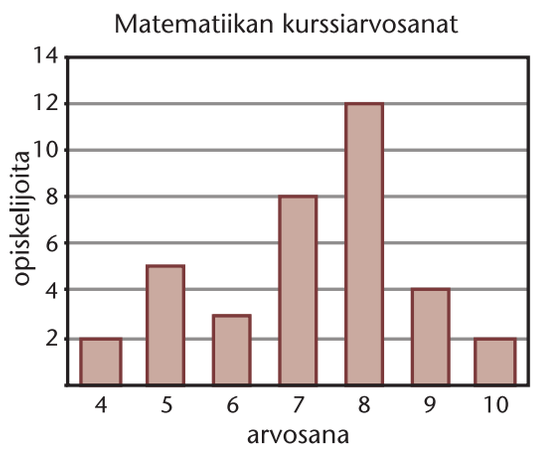
Poissaolopäivät	*f*
5	10
6	6
7	4
8	3
9	1

17. Taulukossa on kunnallisveroprosentit otoksesta, jonka muodostaa 30 Suomen kuntaa. Prosentit on luokiteltu viiteen luokkaan.

Kunnallisveroprosentti	*f*
17,0–17,4	3
17,5–17,9	6
18,0–18,4	5
18,5–18,9	8
19,0–19,4	8

a) Määritä graafisesti likiarvo ala- ja yläkvartiilille.
b) Määritä kvartiiliväli ja laske sen pituus.
c) Kuinka monta prosenttia kvartiiliväli on vaihteluvälistä?

18. Opettaja laati pitämänsä matematiikan kurssin arvosanojen pohjalta tilaston. Määritä kuvaajan avulla aineiston ala- ja yläkvartiilit. Miten ne tulkitaan tässä aineistossa?



Kahden muuttujan yhteisjakauma

Regressiokäyrä

Havaintoaineiston muuttujan arvoista voidaan piirtää erilaisia jakauman kuvaajia. Kuvaajien perusteella voidaan tehdä suuntaa-antavia päätelmiä. Useimmiten tutkimuksissa ollaan kiinnostuneita myös eri muuttujien välisistä suhteista ja riippuvuuksista. Niiden tarkasteluun voidaan käyttää sekä graafisia että laskennallisia menetelmiä.

Kahden muuttujan välistä riippuvuutta voidaan havainnollistaa graafisesti piirtämällä muuttujien **yhteisjakauma**. Yhteisjakaumaa piirrettäessä tulee ensin selvittää, kumpi muuttujista on **selittävä muuttuja** ja kumpi **selitettävä muuttuja**.

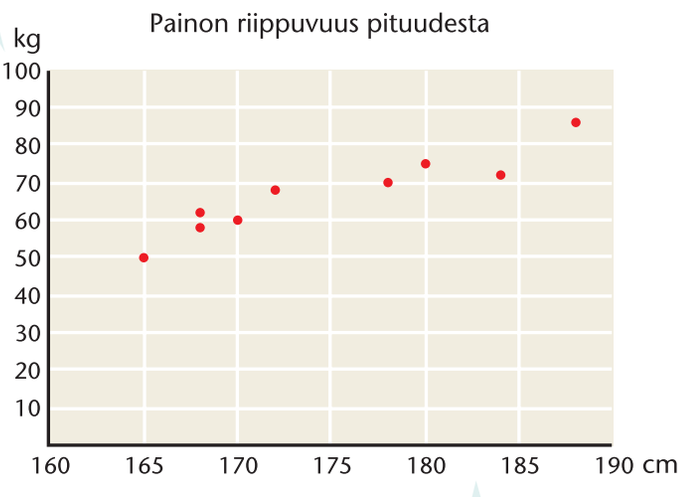
Selittävän muuttujan arvoja kuvataan x-akselilla ja selitettävän muuttujan arvoja y-akselilla. Yhteisjakauma piirretään merkitsemällä koordinaatistoon muuttujien arvojen mukaiset pisteet. Akselien asteikot kannattaa valita siten, että

- pisteet on helppo merkitä koordinaatistoon
- pisteparven muoto saadaan näkyviin.

Tästä syystä asteikkoja ei aina aloiteta nollasta.

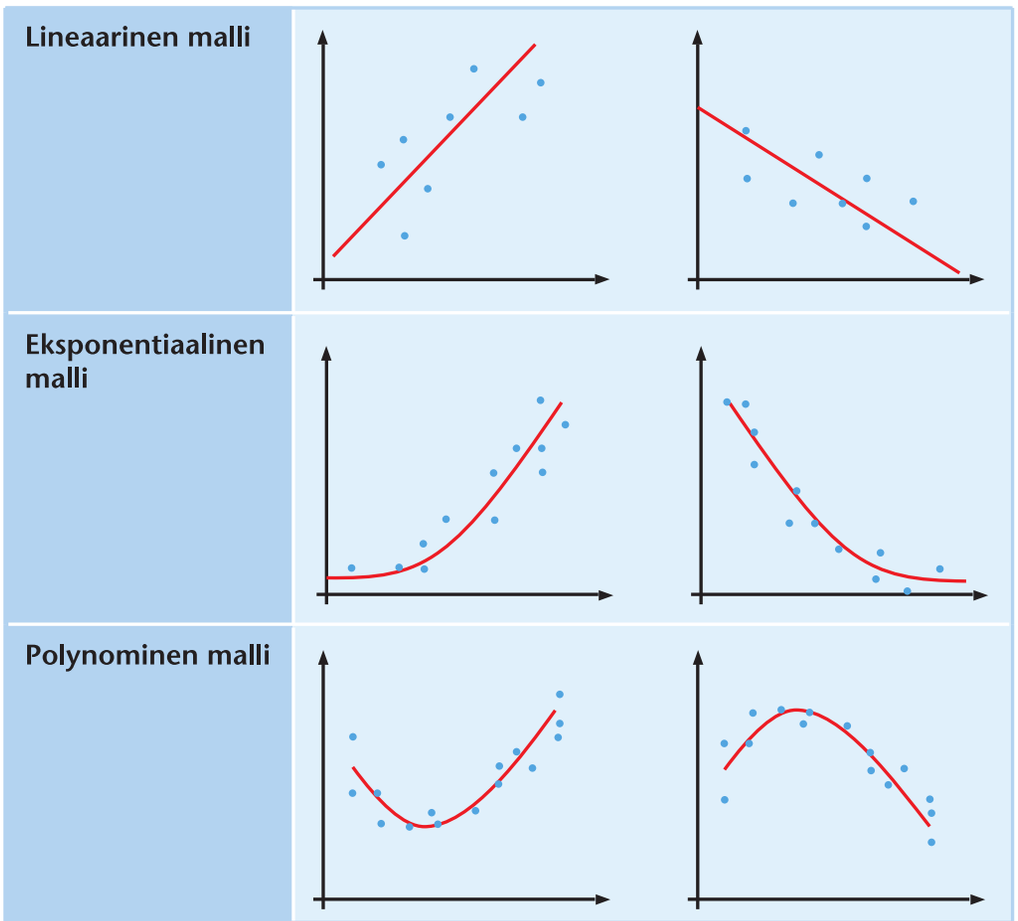
Selitettävä muuttuja: paino

Pituus x	Paino y
165	50
168	62
168	58
170	60
172	68
178	70
180	75
184	72
188	86

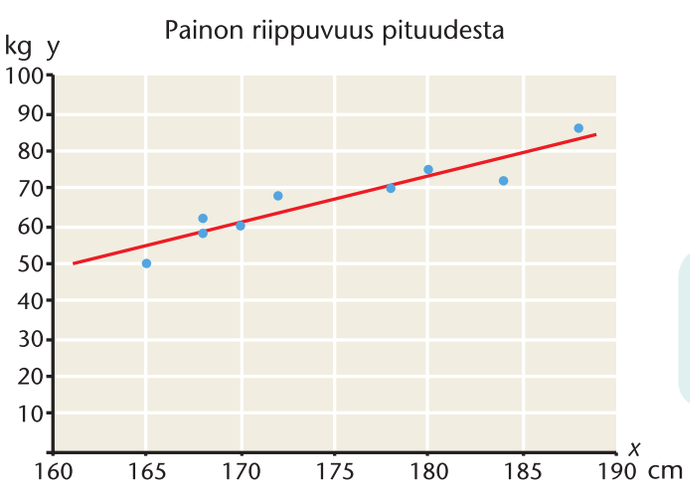


Jos yhteisjakaumaa kuvaavassa hajontakuviossa on säännönmukaisuutta, muuttujien arvot ovat riippuvaisia toisistaan. Tällöin hajontakuviota yritetään usein kuvata jollain matemaattisella mallilla. Tämä voidaan tehdä graafisesti sijoittamalla käyrä pisteparven keskelle.

Pisteparveen sijoitettavan käyrän tulee kuvata pisteparvea mahdollisimman hyvin. Käytännössä tämä tarkoittaa, että käyrän ylä- ja alapuolella on suurin piirtein yhtä monta pisteparven pistettä.



Mallin sovittamisen tarkoituksena on saada selville muuttujien välistä riippuvuutta kuvaava funktio. Tätä funktiota kutsutaan **regressiokäyräksi**. Yleensä malli sovitetaan asianmukaisella tilastollisella tietokoneohjelmalla tai laskimen tilastotoiminnoilla. Mallin sovittamisen jälkeen muuttujien arvoille voidaan laskea **ennusteita** funktion avulla.



Pistejoukkoa kuvaava funktio $y = 1{,}243x - 150{,}38$

Esimerkki 1

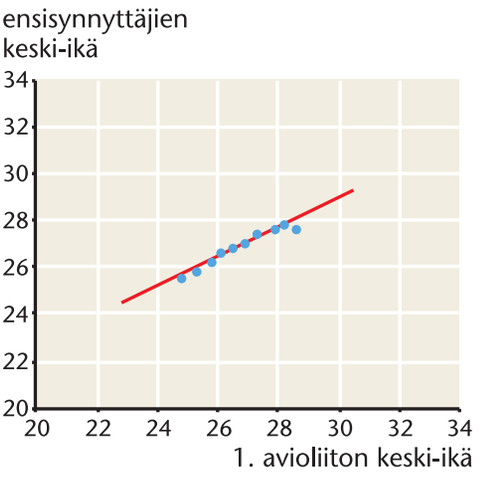
Taulukkoon on koottu vuosien 1982–2000 väliseltä ajalta ensimmäisen avioliiton solmineiden naisten keski-iät ja ensisynnyttäjien keski-iät. Tarkastellaan ensisynnyttäjien keski-iän riippuvuutta 1. avioliiton solmimisen keski-iästä.

a) Piirrä yhteisjakauman kuvaaja ja sovita pistejoukkoon sopiva käyrä. Mikä malli kuvaa tilannetta parhaiten?
b) Määritä käyrän yhtälö.
c) Ennusta mallin avulla minkä ikäinen on ensisynnyttäjä, jos ensimmäinen avioliitto solmitaan 30 vuoden iässä.

Vuosi	1. avioliiton solmimisen keski-ikä	Ensisynnyttäjien keski-ikä
1982	24,8	25,5
1984	25,3	25,8
1986	25,8	26,2
1988	26,1	26,6
1990	26,5	26,8
1992	26,9	27,0
1994	27,3	27,4
1996	27,9	27,6
1998	28,2	27,8
2000	28,6	27,6

Ratkaisu

a) Ensisynnyttäjien keski-ikää halutaan selittää avioliiton solmimisiällä. Merkitään avioliiton solmimisiät x-akselille ja ensisynnyttäjien keski-iät y-akselille.



Yhteisjakauman pisteet näyttävät asettuvan suoralle, joten tilannetta kuvaa parhaiten lineaarinen malli. Suoraa kutsutaan **regressiosuoraksi**.

b) Laskimen tai taulukkolaskentaohjelman avulla saadaan regressiosuoran yhtälöksi

$$y = 10{,}38 + 0{,}6152x$$

c) Jos avioliitto tulevaisuudessa solmitaan 30 vuoden ikäisenä, niin $x = 30$.

Tällöin $y = 10{,}38 + 0{,}6152 \cdot 30 = 28{,}836 \approx 28{,}8$

VASTAUS

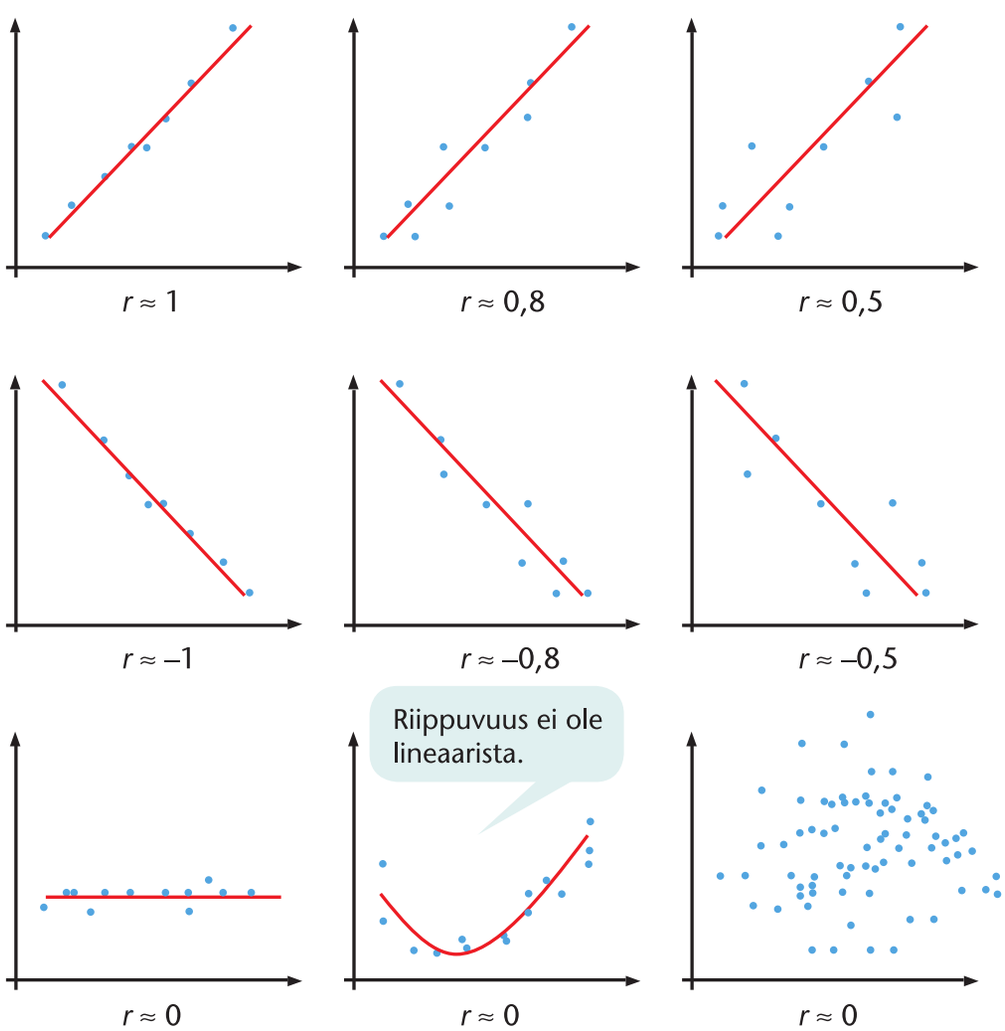
a) lineaarinen malli b) $y = 10{,}38 + 0{,}6152x$
c) keskimäärin 28,8 vuoden ikäisenä

Korrelaatiokerroin

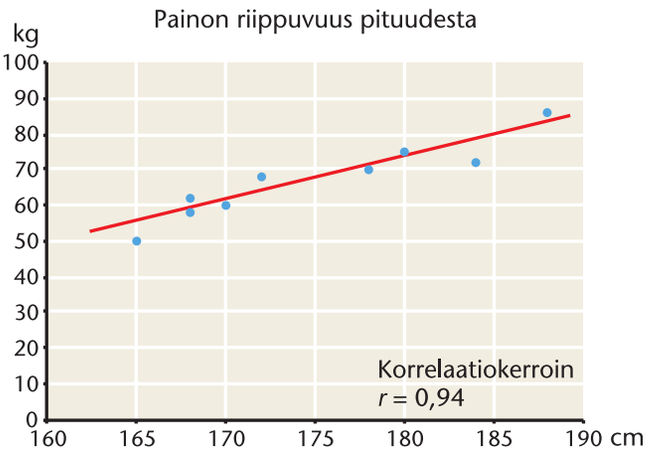
Jos riippuvuus on lineaarista eli pisteet sijoittuvat suoralle, riippuvuuden voimakkuutta voidaan kuvata **korrelaatiokertoimella *r***.

Korrelaatiokerroin

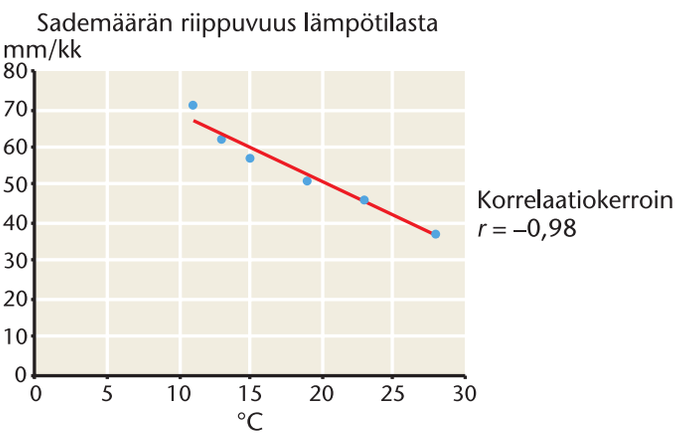
Korrelaatiokerroin r kuvaa kahden muuttujan välisen lineaarisen riippuvuuden voimakkuutta.



Korrelaatiokerroin on yksikötön luku, jonka arvo vaihtelee lukujen −1 ja +1 välillä. Jos korrelaatiokerroin on positiivinen, riippuvuutta kuvaa nouseva suora. Tämä tarkoittaa, että muuttujien arvot muuttuvat samaan suuntaan eli toisen muuttujan arvon kasvaessa toisenkin muuttujan arvo kasvaa.

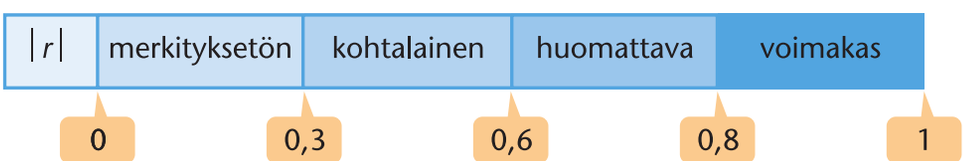


Jos taas korrelaatiokerroin on negatiivinen, riippuvuutta kuvaa laskeva suora. Tämä tarkoittaa, että muuttujien arvot muuttuvat eri suuntiin eli toisen muuttujan arvon kasvaessa toisen muuttujan arvo pienenee.



Korrelaatiokertoimen itseisarvon perusteella voidaan tehdä päätelmiä riippuvuuden voimakkuudesta. Mitä lähempänä korrelaatiokertoimen itseisarvo on lukua 1, sitä voimakkaammasta lineaarisesta riippuvuudesta on kysymys. Jos korrelaatiokerroin on nolla tai hyvin lähellä nollaa, muuttujien välillä ei ole lineaarista riippuvuutta.

Riippuvuuden voimakkuutta voidaan kuvata seuraavalla kaaviolla.



Tehtäviä

19. Erään oppilaan matematiikan kurssin arvosana riippui kurssin oppilasmäärästä seuraavan taulukon mukaisesti:

Oppilasmäärä kurssilla	25	27	20	34	36	32	22	37	28	28	32	31	29	18	26
Kurssiarvosana	9	10	9	8	7	7	8	5	7	7	5	6	8	10	7

Onko oppilasmäärän ja arvosanan välillä riippuvuutta? Jos on, niin mikä malli kuvaa riippuvuutta parhaiten?

20. Lämpötila vaikuttaa veden tiheyteen. Mikä malli kuvaa riippuvuutta parhaiten?

°C	0	1	7	9	15	19	25	29	30
mg/cm³	999,8	999,9	999,9	1000	999,1	998,4	997,1	996	995,7
°C	35	40	45	50	60	70	80	90	100
mg/cm³	994,1	992,2	990,3	988	983,2	977,8	971,8	965	958,4

21. Taulukossa on vuosittaiset työttömyysasteet ja maksamattomien verorästien määrät Suomessa. Onko niiden välillä riippuvuutta? Jos on, millaista?

Vuosi	**Työttömyys aste (%)**	**Ulosottoon lähetetyt (mrd €)**
1990	3,2	0,77
1991	6,6	0,98
1992	11,7	0,93
1993	16,3	1,08
1994	16,6	0,86
1995	15,4	0,81
1996	14,6	0,78
1997	12,7	1,32
1998	11,4	1,10
1999	10,2	1,00
2000	9,8	1,32

22. Tietyt atomit voivat muodostaa keskenään kovalenttisen sidoksen. Sidoksen energian ja muodostuvan sidoksen pituuden välillä on voimakas riippuvuus. Tutki tätä riippuvuutta koordinaatistossa. Sovita pisteparveen regressiokäyrä ja arvioi sen avulla, mikä on sidoksen energia, kun sen pituus on
a) 300 pikometriä b) 50 pikometriä.

Atomipari	H–H	O–H	N–H	C–H	C–H	S–H	Br–H	P–H	Si–H	C–C	I–H	Cl–Cl	Br–Br	Si–Si	I–I
Pituus (pm)	74	96	101	109	128	135	141	142	146	154	160	199	228	235	267
Energia (kJ)	436	463	388	412	431	338	366	322	318	348	299	242	193	176	151

23. Putoamiskiihtyvyys merenpinnan tasolla vaihtelee maapallon muodon mukaan. Putoamiskiihtyvyys riippuu siitä, kuinka kaukana päiväntasaajasta ollaan. Tarkastele riippuvuutta kuvaajan avulla. Sovita pistejoukkoon regressiokäyrä ja päättele sen avulla, mikä on putoamiskiihtyvyys päiväntasaajalla.

Leveyspiiri	Putoamiskiihtyvyys (m/s^2)
10	9,7820
20	9,7864
30	9,7933
40	9,8017
50	9,8107
60	9,8197
70	9,8261
80	9,8306
90	9,8322

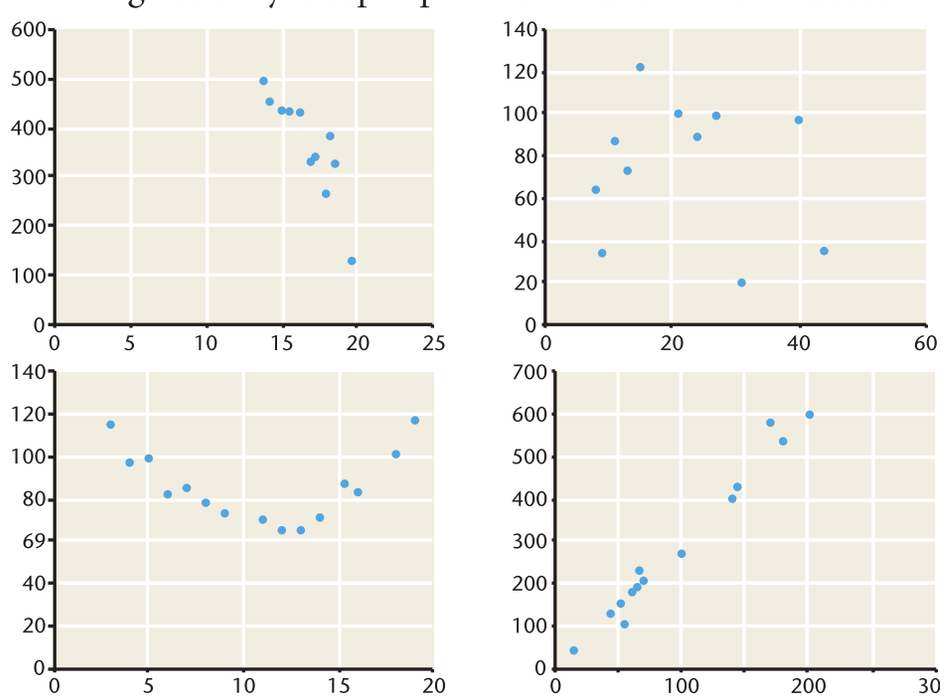
24. Taulukossa on tietoja bruttokansantuotteen kehityksestä (€/asukas) ja imeväiskuolleisuudesta (alle 1-vuotiaina kuolleet 1000 syntynyttä lasta kohti). Tarkastele BKT:n vaikutusta imeväiskuolleisuuteen sovittamalla regressiokäyrä pistejoukkoon.
a) Määritä regressiokäyrän yhtälö.
b) Arvioi mallin avulla imeväiskuolleisuus vuonna 2002, jolloin BKT oli 26 863 € asukasta kohti.

	BKT €/asukas	Imeväiskuolleisuus
1982	8 633	6,0
1984	10 663	6,5
1986	12 302	5,8
1988	15 079	6,1
1990	17 665	5,6
1992	16 265	5,2
1994	17 271	4,7
1996	19 246	4,0
1998	22 572	4,2
2000	25 144	3,8

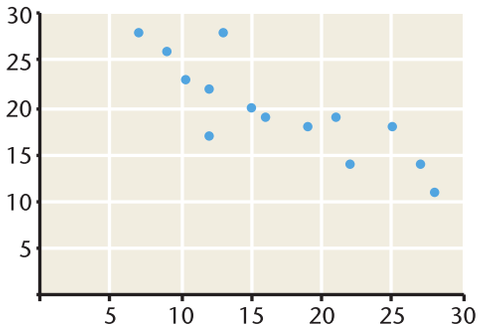
25. a) Missä seuraavista kuvista on havaittavissa riippuvuutta tutkittavien suureiden välillä?
b) Yhdistä oikea kuva ja korrelaatiokerroin.

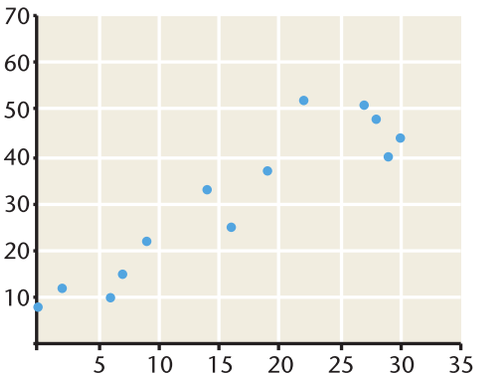
$r \approx 0{,}97$	$r \approx 0$	$r \approx -0{,}15$	$r \approx -0{,}87$

c) Mikä regressiokäyrä sopisi parhaiten kuhunkin tilanteeseen?

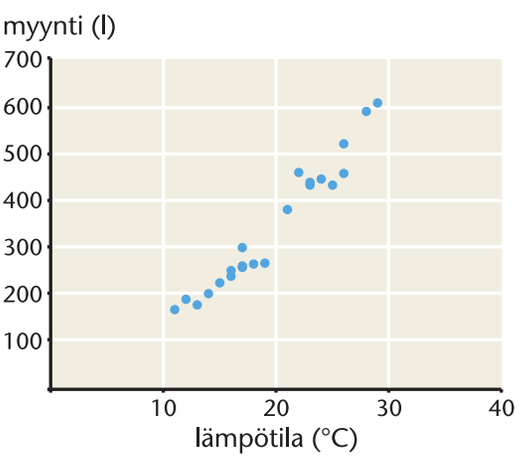


26. Sijoita väittämä oikeaan kuvaan *A* tai *B*.
    a) Suureilla on melko voimakas positiivinen korrelaatio.
    b) Korrelaatiokertoimen arvo on noin –0,84.
    c) Regressiokäyrä on nouseva suora.
    d) Regressiokäyrä on laskeva suora.
    e) Suureilla on voimakas negatiivinen korrelaatio.
    f) Korrelaatiokertoimen arvo on noin 0,93.





27. Kuvaaja esittää valintamyymälän jäätelönmyyntiä (l) kesäaikaan. Päivittäiset lämpötilat ja myynnit muodostavat pisteparven, joka on sijoitettu koordinaatistoon.



    a) Arvioi korrelaatiokertoimen *r* arvoa sijoittamalla se yhteen seuraavista luokista:

$-1 < r < -0{,}5$	$-0{,}5 < r < 0$	$0 < r < 0{,}5$	$0{,}5 < r < 1$

    b) Tulkitse korrelaatiokerrointa.

28. Tilastokeskuksen työvoimatutkimuksen mukaan palkansaajia ja yrittäjiä oli vuosina 1960–1991 tuhansina henkilöinä seuraavat määrät. Havainnollista tilastotiedot diagrammina samassa kuviossa. Sovita kuvioon havaintopisteisiin mahdollisimman hyvin sopivat (regressio)suorat ja esitä tämän aineiston pohjalta ennuste, kuinka paljon palkansaajia ja yrittäjiä on Suomessa vuonna 1994. [K1994, 10c]

Vuosi	Palkansaajia	Yrittäjiä
1960	1 340	757
1970	1 626	500
1980	1 930	379
1985	2 077	360
1990	2 108	359
1991	2 000	339

29. Taulukossa on Sydneyn olympialaisten kymmenottelun kymmenen parhaan urheilijan pistetulos 100 metrin juoksussa ja kuulantyönnössä sekä ottelun lopulliset kokonaispistemäärät. Tutki, miten 100 metrin juoksun ja kuulantyönnön yhteistuloksen perusteella voidaan ennustaa lopullista menestymistä kymmenottelussa. Laske tätä tarkoitusta varten 100 metrin juoksun sekä kuulantyönnön pistesumman ja ottelun kokonaispistemäärän korrelaatiokerroin sekä piirrä jakaumasta korrelaatiodiagrammi. Sovita kuvioon regressiosuora. [K2001, 15]

	100 m juoksu	Kuulantyöntö	Kokonaispistemäärä
1. Erki Nool EST	933	796	8641
2. Roman Sebrle CZE	878	803	8606
3. Chris Huffins USA	980	806	8595
4. Dean Macey GBR	903	766	8567
5. Tom Pappas USA	901	782	8425
6. Tomas Dvorak CZE	881	846	8385
7. Frank Busemann GER	881	760	8351
8. Attila Zsivoczky HUN	838	787	8277
9. Stefan Schmid GER	874	731	8206
10. Henrik Dagård SWE	897	788	8178

Eduard Hämäläisen tulos 100 metrin juoksussa oli 858 pistettä ja kuulantyönnössä 732 pistettä. Jos edellä saatua regressioanalyysin tulosta käytetään ennusteena, niin mikä olisi ollut Hämäläisen odotettavissa oleva lopullinen kokonaispistemäärä?

Tehtävien vastaukset

1.1 Tilastotieteen peruskäsitteitä

1. a) 13,6 % b) 639 396
   c) 27,6 %

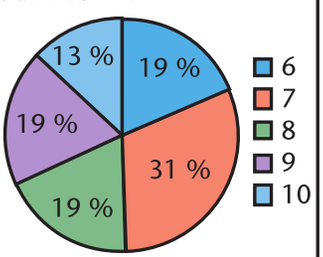
2. a) 62 % b) 86 %

3. a)

Arvosana	*f*	*f* %
6	3	18,75 %
7	5	31,25 %
8	3	18,75 %
9	3	18,75 %
10	2	12,50 %
Yhteensä	16	100,0 %

Moodi Mo = 7

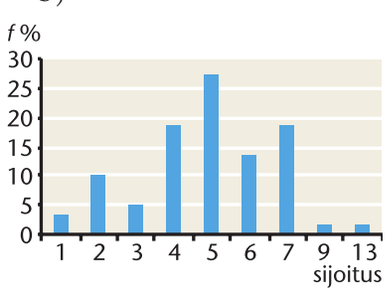
b)



4. a)

Sijoitus	*f*	*f* %
1	2	3,4 %
2	6	10,2 %
3	3	5,1 %
4	11	18,6 %
5	16	27,1 %
6	8	13,6 %
7	11	18,6 %
9	1	1,7 %
13	1	1,7 %

b)



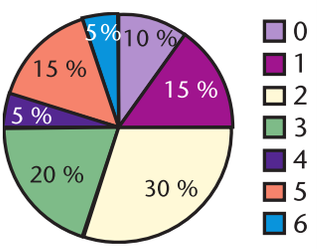
5. a) 166 160 €
   b) 22 320 €
   c) 14 880 €

6. a)

Liikunta-pv/vko	*f*	*f* %
0	2	10 %
1	3	15 %
2	6	30 %
3	4	20 %
4	1	5 %
5	3	15 %
6	1	5 %

b) Mo = 2

c)



7. a)

Paino (kg)	Luokkakeskus	Luokkavälin pituus	Todellinen alaraja	Todellinen yläraja
50–54	52	5	49,5	54,5
55–59	57	5	54,5	59,5
60–64	62	5	59,5	64,5

b)

Juoksuaika (min)	Luokkakeskus	Luokkavälin pituus	Todellinen alaraja	Todellinen yläraja
24–29	26,5	6	23,5	29,5
30–35	32,5	6	29,5	35,5
36–41	38,5	6	35,5	41,5

8. Esimerkiksi

a)

Korkeusero (m)	Todellinen alaraja (m)	Todellinen yläraja (m)
50–99	49,5	99,5
100–149	99,5	149,5
150–199	149,5	199,5
200–249	199,5	249,5
250–299	249,5	299,5
300–349	299,5	349,5

b)

Korkeusero (m)	Luokkakeskus (m)
50–99	74,5
100–149	124,5
150–199	174,5
200–249	224,5
250–299	274,5
300–349	324,5

c) 50 m

9. Esimerkiksi

Luokka (km/h)	*f*	Luokkakeskus (km/h)
61–70	1	65,5
71–80	2	75,5
81–90	3	85,5
91–100	2	95,5
101–110	4	105,5
111–120	2	115,5

10. Esimerkiksi:

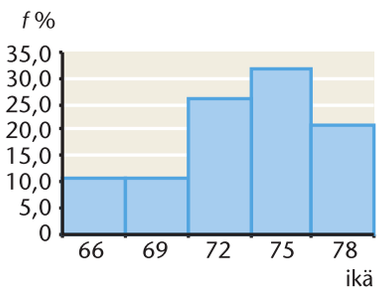
a)

	f	Luokkakeskus
1–3	9	2
4–6	6	5
7–9	3	8
10–12	2	11

b) Todelliset luokkarajat ovat samat kuin taulukossa ilmoitetut luokkarajat.

11.

Ikä	*f*	*f* %	Luokka-keskus
65–67	2	10,5 %	66
68–70	2	10,5 %	69
71–73	5	26,3 %	72
74–76	6	31,6 %	75
77–79	4	21,1 %	78
Yhteensä	19	100,0 %	



12.

Luokka (%)	*f*	Luokka-keskus (%)
0,1–1,0	3	0,55
1,1–2,0	1	1,55
2,1–3,0	13	2,55
3,1–4,0	10	3,55
4,1–5,0	1	4,55
5,1–6,0	2	5,55

Mo = 2,55 %

13.

Äänestys-prosentti	*f*	*f* %	Luokka-keskus (%)
56–60	6	15,0 %	58
61–65	9	22,5 %	63
66–70	16	40,0 %	68
71–75	7	17,5 %	73
76–80	2	5,0 %	78

14. a) 0,6 b) 0,005

15. a) 0,123 b) 0,877

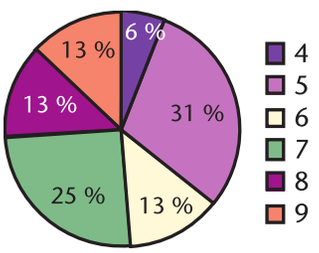
16. $\frac{41}{155} \approx 0{,}265$

17. a)

Lyönnit (lkm)	*f*	*f* %
4	1	6,25 %
5	5	31,25 %
6	2	12,5 %
7	4	25 %
8	2	12,5 %
9	2	12,5 %

Mo = 5

b)

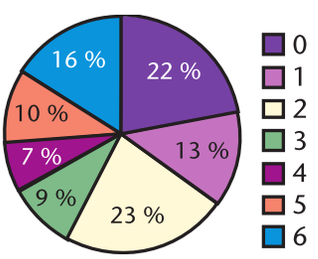


18. Esimerkiksi

Leijonat (lkm)	*f*	*f* %
5–8	7	32 %
9–12	7	32 %
13–16	5	23 %
17–20	3	14 %

Mo = 6,5 tai Mo = 10,5

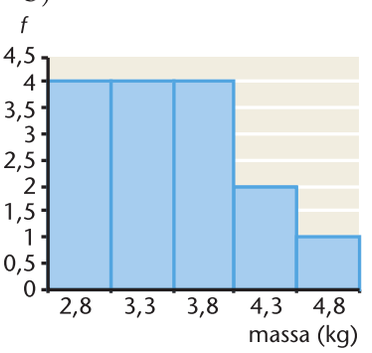
19.



20. a)

Massa (kg)	*f*	*f* %	Luokka-keskus (kg)
2,6–3,0	4	27 %	2,8
3,1–3,5	4	27 %	3,3
3,6–4,0	4	27 %	3,8
4,1–4,5	2	13 %	4,3
4,6–5,0	1	7 %	4,8

b)



21. a) 0,079 b) 0,87

1.2 Summafrekvenssit

22. a) 27 % b) 11 c) 88 %

23. a)

Tietokone (lkm)	*f*	*sf*	*sf* %
0	8	8	1 %
1	191	199	23 %
2	462	661	77 %
3	136	797	93 %
4	58	855	99 %
5	6	861	100

b) 93 % c) 7 %

24. a) 46 % b) 20 %
c) 98 %

25. a) 11 oppilasta
b) 6 %
c) 67 %
d) 14 oppilasta
e) 6 oppilasta
f) 62 %
g) 18 oppilasta eli kaikki

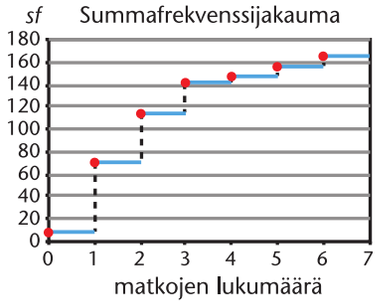
26. a) 7 % b) 14 %

27.

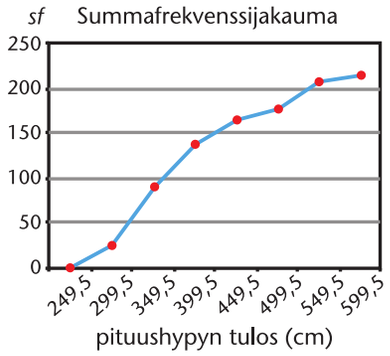
f	*f* %	*sf*	*sf* %
2	10 %	2	10 %
4	20 %	6	30 %
10	50 %	16	80 %
4	20 %	20	100 %

28. A–E, B–F, C–D

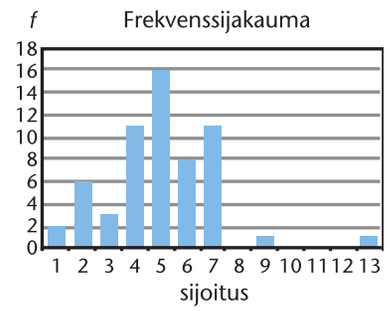
29. a)



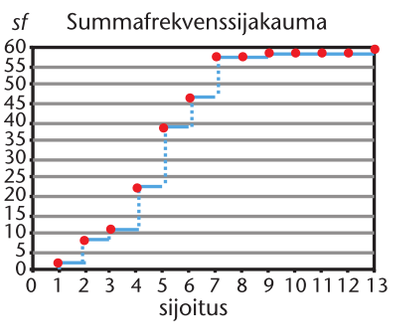
b)



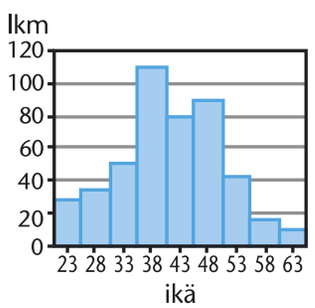
30. a)



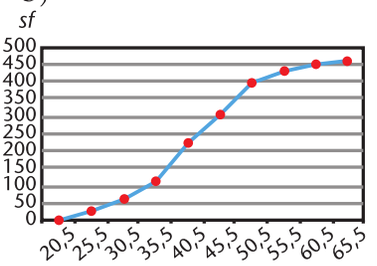
b)



31. a)



b)



32. Mo = 9, Md = 8,5

33. a) 7 eli heinäkuu, tällöin yöpyy eniten turisteja
b) Md = 7
c) 470 383 yöpymistä

34. a) Mo = 2, Md = 3
b) Mo = 2249,50 € ja Mo = 3249,50 €, Md = 2749,50 €

35. Mo = 349,5 mökkiä ja Mo = 749,5 mökkiä, Md = 449,5 mökkiä
Kunnissa yleisimmin 349,5 mökkiä tai 749,5 mökkiä.
Puolet kyseisistä kunnista sellaisia, joilla mökkien määrä on korkeintaan 449,5 mökkiä.

36. a) 62 kg b) 2,7 km

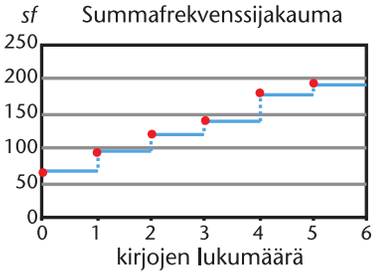
37. a) Mediaaniluokka: 75–77 kurssia eli puolet suorittaa tämän määrän kursseja ja puolet enemmän.
b) Moodiluokka: 75–77 eli tämä on yleisin suoritettu kurssimäärä.

38. 42 €

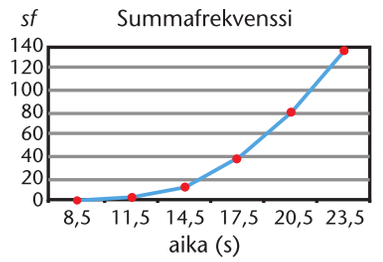
39. a) Md = 74,5 min
b) Md ≈ 72 min

40. a) 49 peliä
b) 51 %
c) 41 %

41. a)



b)



42. Mediaani on klo 14.00. Klo 14 mennessä on käynyt puolet päivän asiakkaista.

43. a) Md = 464,5 min
b) Md ≈ 457 min

1.3 Keskiarvo

44. a) 604 488 b) 484 487

45. a) 9,94 % b) 10,0 %

46. Anssin keskiarvoksi tulee 7,3 eli ei voita.

47. a) Mo = 0 h, Md = 2 h
b) 2,6 h

48. 2,8 %

49. Mo = 15–64, Md = 15–64, $\bar{x} = 39{,}1$

50. Esimerkki luokittelusta:

Osuus (%)	f	Luokka-keskus (%)
26–30	1	28
31–35	4	33
36–40	3	38
41–45	6	43
46–50	3	48
51–55	1	53

Näin luokiteltuna
Md = 43 % (41–45),
Mo = 43 % (41–45),
$\bar{x} \approx 40{,}5\,\%$

51. a) 4,10 g b) 5,50 €

52. a) 10 b) 4

53. –1,8 ja 0,8

54. 30 milj. €

55. 8,6

56. 6,89 dm^3/100 km

57. 78,993275

58. a) Mo = 6, Md = 5
b) 5,4 kpl

59. $\bar{x} \approx 7{,}5$ (7,49)

60. Esimerkki luokittelusta:

Pisteet	f	Luokka-keskus
0–9	18	4,75
10–19	5	14,5
20–29	2	24,5
30–39	4	34,5
40–49	1	44,5

Näin luokiteltuna
Mo = 0–9, Md = 0–9,
$\bar{x} \approx 13$

61. 63 kg

62. –

1.4 Keskihajonta

63. 13 dB

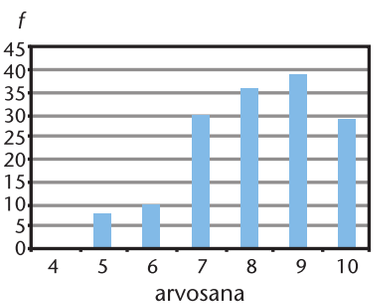
64. a) 165 elokuvaa
b) 15 elokuvaa

65. a) Matti b) Toni

66. 5 vuoden: $\bar{x} \approx 5{,}29\,\%$,
$s \approx 1{,}34\,\%$
10 vuoden: $\bar{x} \approx 5{,}98\,\%$,
$s \approx 1{,}48\,\%$
10 vuoden obligaatioissa on parempi tuotto, mutta tuotto vaihtelee enemmän vuodesta riippuen.

67. $\bar{x} \approx 3059$ m, $s \approx 1600$ m

68. $\bar{x} \approx 8{,}15$ ja $s \approx 1{,}39$



69. Helsinki: $\bar{x} \approx 17{,}8\,°C$
ja $s \approx 1{,}74\,°C$
Md = 17,85 °C
Sodankylä: $\bar{x} \approx 14{,}9\,°C$
ja $s \approx 0{,}992\,°C$
Md = 15,2 °C

70. a) $s \approx 9{,}85$ s
b) Eivät, sillä molemmat arvot ovat alle kahden keskihajonnan päässä keskiarvosta.
($\bar{x} = 137{,}476\ldots$ (s))

71. $\bar{x} = 99$, $s \approx 22{,}5$
Vuonna 1991 poikkeaa, vuonna 2001 ei poikkea.

72. a)

Massa (g)	f	Luokka-keskus (g)
200–249	2	224,5
250–299	4	274,5
300–349	3	324,5
350–399	5	374,5
400–449	2	424,5

b) $\bar{x} \approx 328$ g ja $s \approx 64{,}5$ g
c) Poikkeaa, arvo on melkein kolmen keskihajonnan päässä keskiarvosta.

73. $s \approx 3{,}7$ m/s

74. a) Elinan b) Elinan

75. Keskiarvo on 3,10 ja keskihajonta 1,56.

76. a) $\bar{x} \approx 148\ m^2$ ja
$s \approx 54{,}4\ m^2$
b) Ei, ala on 1,4 keskihajonnan päässä keskiarvosta.

77. $s \approx 9{,}29$

1.5 Normaalijakauma

78. Vilman koira

79. 105 vuotta

80. 6,2

81. a) 0 b) –1 c) 1

82. a) 0,60 b) 0,91
c) 0,085

83. 59 %

84. a) 0,89 b) 0,11

85. a) 0,88 b) 0,12

86. a) 0,021
b) 5 oppilasta

87. a) 0,89 b) 0,040

88. a) $z_{16,7} \approx 0{,}49$, $z_{20,5} \approx 1{,}57$
b) 0,058
c) 0,25

89. 0,52

90. a) 90,8 % b) 9,2 %
c) 65,7 % d) 25,1 %

91. 0,31

92. a) 0,87 b) 0,13

93. a) 0,048 b) 0,20

94. 0,68

95. a) 0,9 m b) 86 cm

96. 23,8 cm

97. 7,3 min

98. enintään 0,07 mm

99. 47 pistettä

100. 17 cm

101. a) 0,67 b) 0,023
c) 0,68

102. a) 0,34 b) 0,66
c) 0,34

103. a) 0,076 b) 0,076
c) 0,85

104. 4,8 g

105. a) 48 pistettä
b) 89,6 %

2.1 Malleja todennäköisyyden määrittämiseksi

106. a) 0,54 b) 0,29

107. a) 0,019 b) 0,0064
c) 0,97

108. a) 0,32 b) 0,39

109. a) $\frac{3}{10}$ b) $\frac{1}{2}$

110. a) 0,4 b) 0,7
c) 0,6

111. a) $\frac{1}{8}$ b) $\frac{1}{2}$

112. 0,1

113. a) $\frac{1}{6} \approx 0{,}17$ b) $\frac{5}{18} \approx 0{,}28$
c) $\frac{5}{9} \approx 0{,}56$

114. 7

115. 0,1

116. a) 0,0028 b) 0,069
c) 0

117. 0,69

118. 0,25

119. 0,053

120. a) 0,17 b) 0,25
c) 0,35

121. 0,04

122. a) 0,5 b) 0,125
c) 0,01

123. a) 0,51 b) 0,50

124. a) $\frac{1}{3}$ b) $\frac{1}{8}$
c) $\frac{1}{24}$

125. a) $\frac{1}{7} \approx 0{,}14$ b) $\frac{5}{28} \approx 0{,}18$
c) $\frac{1}{28} \approx 0{,}036$

126. a) $\frac{1}{144} \approx 0{,}0069$
b) $\frac{25}{144} \approx 0{,}17$

127. a) $\frac{3}{10}$ b) $\frac{1}{10}$

128. a) $\frac{4}{25}$ b) $\frac{12}{25}$

2.2 Komplementti ja kertolaskusääntö

129. 0,12

130. $\frac{1}{8} = 0{,}125$

131. a) 0,045 b) 0,55

132. a) 0,00077 b) 0,012

133. a) 0,6 b) $6 \cdot 10^{-9}$

134. a) 0,0001 b) 0,34

135. 0,74

136. a) 0,35 b) 0,65

137. a) 0,26 b) 0,99994

138. a) 0,0024 b) 0,074

139. 0,95; 0,86; 0,14

140. a) 0,19 b) 0,81
c) 0,994

141. 0,091

142. 0,00018

143. a) 0,75 b) 0,78

144. a) 0,12 b) 0,021
c) 0,98 d) 0,88

145. 7 kappaletta

146. 0,58

147. 4 pataa

148. 16 henkilöä

149. 0,18

150. a) 0,021 b) 0,0000027

151. 0,47

152. a) 0,064 b) 0,936
c) 0,718

153. a) 0,000035
b) $3{,}7 \cdot 10^{-6}$

154. 21 poikaa

2.3 Yhteenlaskusääntö

155. a, b, c ja e

156. a) 0,50 b) 0,31

157. 0,41

158. a) 0,475 b) 0,53
c) 0,65

159. a) 0,0525 b) 0,605

160. 0,018

161. $\frac{3}{8} \approx 0{,}38$

162. a) $\frac{1}{21} \approx 0{,}048$

b) $\frac{10}{21} \approx 0{,}48$

163. a) 0,21 b) 0,79

164. a) 0,50 b) 0,40

165. a) 0,833 b) 0,164
c) 0,003

166. a) $\frac{1}{10}$ b) $\frac{1}{2}$

167. 0,84

168. 0,1525 ≈ 0,15

169. 0,38

170. a) 0,14 b) 0,047
c) 0,14

171. $\frac{1}{9} \approx 0{,}11$

172. a) 0,0225 ≈ 0,023
b) 0,1075 ≈ 0,11

2.4 Tuloperiaate ja kombinaatio

173. 60

174. 24 389 000

175. 21 000

176. a) 1 048 576
b) $9{,}5 \cdot 10^{-7}$

177. 63

178. Ainakin yksi on punainen todennäköisyydellä 0,9161. Kaikki ovat samanvärisiä todennäköisyydellä 0,0842.

179. $2{,}7 \cdot 10^{32}$

180. $3{,}97 \cdot 10^{-4}$

181. 36

182. 0,33

183. 0,042

184. a) 35 b) 1716

185. 80 730

186. 455

187. a) 2 598 960
b) 65 780
c) 48

188. $\frac{1}{20} = 0{,}05$

189. 180

190. 0,41

191. 4 oikein 173 600
5 oikein 10416
6 oikein 224

192. 0,054

193. a) 0,017 b) 0,34

194. a) 9
b) 1120 tavalla

195. a) 5040 b) 144
c) 240 d) 720

196. a) 3276 b) 98 280

197. a) $\frac{1}{512} \approx 0{,}00195$
b) $\frac{27}{64} \approx 0{,}422$

198. a) 0,23 b) 0,0032

2.5 Binomitodennäköisyys

199. a) 0,17 b) 0,55

200. –

201. 0,12

202. a) 0,047 b) 0,14

203. a) 0,17 b) 0,0017

204. 0,080

205. 0,053

206. a) 0,41 b) 0,84

207. a) 0,25 b) 0,96
c) 0,14

208. a) 0,23 b) 0,0028

209. a) 0,050 b) 0,42

210. a) 0,15 b) 0,20

211. a) 0,16 b) 0,20

212. 0,20

213. a) 0,057 b) 0,34
c) 0,29

214. a) 0,23 b) 0,97

215. a) 0,95 b) 0,000041

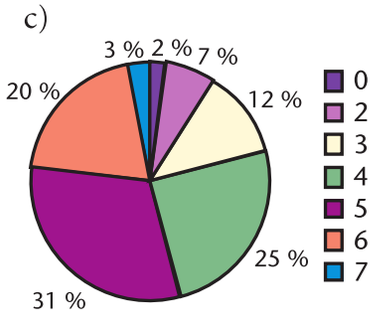
Kertausosa

1. a)

Äänet (lkm)	*f*	*f* %
0	1	1,7 %
2	4	6,8 %
3	7	11,9 %
4	15	25,4 %
5	18	30,5 %
6	12	20,3 %
7	2	3,4 %

b) Mo = 5

c)

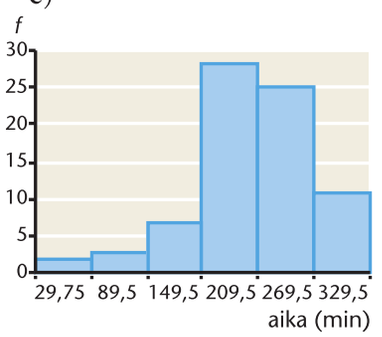


2. a)

Aika (min)	*f*	*f* %
0–59	2	2,6 %
60–119	3	3,9 %
120–179	7	9,2 %
180–239	28	36,8 %
240–299	25	32,9 %
300–359	11	14,5 %

b) Moodiluokka = 180–239 (min)
Mo = 209,5 min

c)



3.

Pituus (cm)	*f*	*f* %	Luokka-keskus
155–159	1	1,3 %	157 cm
160–164	0	0,0 %	162 cm
165–169	9	11,8 %	167 cm
170–174	5	6,6 %	172 cm
175–179	2	2,6 %	177 cm
180–184	4	5,3 %	182 cm

4. a) $\frac{1}{5}=0{,}2$ b) $\frac{7}{10}=0{,}7$
c) $\frac{3}{10}=0{,}3$

5. a)

Arvo-sana	*f*	*sf*	*sf* %
I	437	437	4,5 %
A	829	1266	13,0 %
B	1863	3129	32,2 %
C	2388	5517	56,8 %
M	2021	7538	77,6 %
E	1653	9191	94,6 %
L	528	9719	100,0 %

b) 56,8 % c) 43,2 %

6.

Ikä	*sf* %
20–24	24,3 %
25–29	45,8 %
30–34	64,1 %
35–39	80,8 %
40–44	100,0 %

Mediaaniluokka 30–34

7. 202 cm^3

8. 4,8

9. 27,7

10. 12,50 €

11. $-4\frac{2}{3}, 6\frac{2}{3}, 9\frac{2}{3}$ ja $20\frac{1}{3}$

12. Mittari B on joka lämpötilassa tarkempi. Virheiden keskiarvot: A 0,03 ja B 0,4. Virheiden itseisarvojen keskiarvo olisi parempi tunnusluku.

13. $\bar{x} \approx 7{,}0$ ja $s \approx 1{,}6$

14. a)

Ala (ha)	*f*
10–29	4
30–49	3
50–69	2
70–89	4
90–109	1
110–129	1

b) $\bar{x} \approx 54{,}8$ ha ja $s \approx 33{,}4$ ha
c) ei poikkea merkitsevästi

15. a) 3004 € – 4119 €
b) ei poikkea

16. a) 0,16 b) 0,90

17. Bertta, P(tulos yli 82 lukiossa A) ≈ 14 %, P(tulos välillä [72, 80] lukiossa B) ≈ 38 %

18. 21 700 km (vastaukseksi hyväksyttiin myös 21 000 km tai 22 000 km)

19. korkeintaan 3,0 minuuttia

20. 42 tikkua

21. a) 0,38 b) 0,27
c) 0,24

22. $\frac{1}{8}$

23. a) $\frac{2}{5}$ b) $\frac{3}{10}$

24. a) $\frac{1}{7}$ b) $\frac{1}{12}$
c) $\frac{1}{6}$

25. a) 0,50 b) 0,16

26. 0,00066

27. $\frac{5}{12} \approx 0{,}42$ ja $\frac{1}{12} \approx 0{,}083$

28. a) 0,058 b) 0,94

29. a) 0,82 b) 0,9998

30. a) 0,00013 b) 0,13

31. 0,96

32. 0,59

33. a) 0,055 b) 0,0018
c) 0,11

34. 7 kpl

35. Silmälukujen todennäköisyydet:
$\frac{1}{21}, \frac{2}{21}, \frac{3}{21}, \frac{4}{21}, \frac{5}{21}, \frac{6}{21}$
Molemmilla heitoilla saadaan kuutonen todennäköisyydellä 0,082.

36. a) $\frac{5}{33} \approx 0{,}15$ b) $\frac{19}{66} \approx 0{,}29$

37. a) 0,76 b) 0,62

38. a) 0,048 b) 0,38

39. a) 36 b) 518400

40. a) 495 b) 0,018

41. a) 0,50 b) 0,42
c) 0,030

42. a) 0,0011 b) 0,000021

43. 0,11

44. Ei, tapahtuman todennäköisyys on 0,26.

45. a) 0,019 b) 0,98
c) $1{,}6 \cdot 10^{-7}$

Harjoituskoe 1

1. a) 39
b) Mo = 1, Md = 3,
$\bar{x} \approx 3{,}33$
c) $s \approx 1{,}84$

2. a) 0,028 b) 0,20

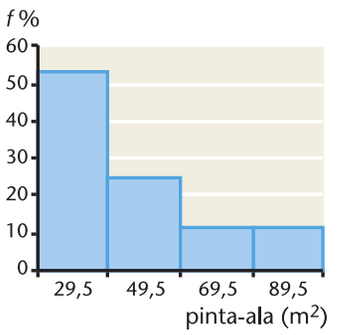
3. 0,21

4. a) 28 561 b) 0,50

5. a)

Pinta-ala (m^2)	f	f %	sf	sf %
20–39	133 823	53 %	133 823	53 %
40–59	61 571	25 %	195 394	78 %
60–79	27 967	11 %	223 361	89 %
80–99	27 105	11 %	250 466	100 %

b) 78 %
c) 20–39 (m^2)
d)



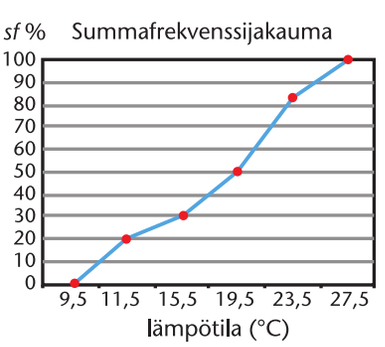
6. a) 0,28 b) 0,65

Harjoituskoe 2

1. a) 45 %
b) $\bar{x} \approx 6{,}1$, Mo = 8 tai
Mo = 5, Md = 6
c) $s \approx 1{,}99$
d) ei

2.

Lämpötila (°C)	f	sf	f %	sf %
10–13	6	6	20 %	20 %
14–17	4	10	13 %	33 %
18–21	5	15	17 %	50 %
22–25	10	25	33 %	83 %
26–29	5	30	17 %	100 %



3. a) 0,71 b) 0,97

4. 0,97

5. a) 23 % b) 15 %

6. a) 120 b) 10
c) 0,4

Harjoituskoe 3

1. Mo = 172 cm,
Md = 167 cm,
$\bar{x} \approx 166$ cm, $s \approx 7{,}40$ cm

2. a) 0,20 b) 0,86
c) 0,36

3. a) $\frac{1}{5}$
b) $\left(\frac{1}{5}\right)^{26} \approx 6{,}7 \cdot 10^{-19}$
c) $\frac{5}{52}$

4. a) 0,28 b) 0,99

5. 33 %

6. 205 cm^3

Ekstrat

1. a) Valitaan satunnaisesti lukiolaisia eri kouluista.
b) Kootaan aineisto vain oman koulun oppilaista.
c) Kootaan aineistoon tiedot kaikilta lukiolaisilta.

2. a) otanta
b) otanta (kokonais)
c) kokonais
d) otanta

3. a) kokonais
b) otanta
c) otanta

4. –

5. –

6. –

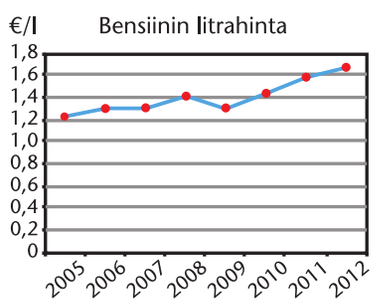
7. a) jatkuva b) diskreetti
c) jatkuva d) diskreetti

8. a) Kvantitatiivisia: pinta-ala, välikulu, verojen osuus
Kvalitatiivisia: valtiomuoto
b) Jatkuvia: pinta-ala, verojen osuus
Diskreetttejä: väkiluku, valtiomuoto
c) Suhdelukuasteikko: pinta-ala, väkiluku, verojen osuus
Laatueroasteikko: valtiomuoto

9. a) suhdeluku
b) järjestys
c) laatuero
d) suhdeluku

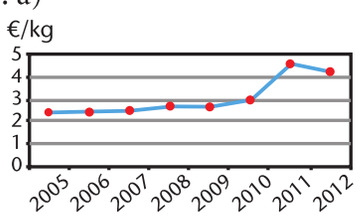
10. Laatueroasteikko: merkki, väri
Välimatka-asteikko: rekisteröintivuosi
Suhdelukuasteikko: ajetut kilometrit, hinta-arvio

11.

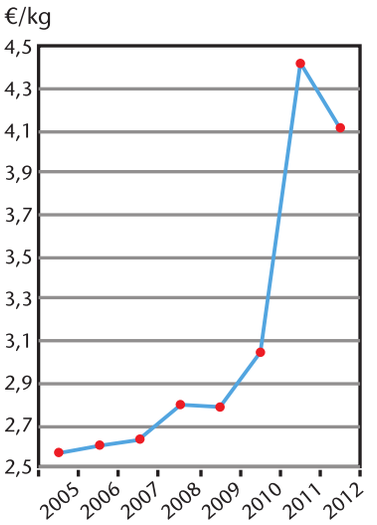


12. –

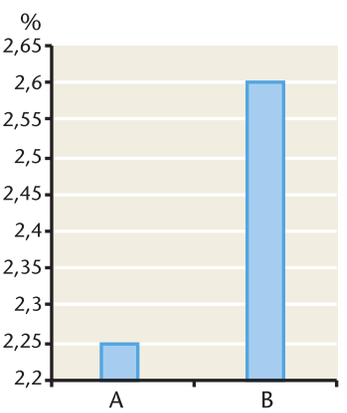
13. a)



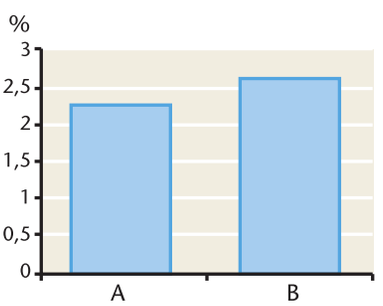
b)



14. a) Eron tulisi näyttää mahdollisimman suurelta.



b) Eron tulisi näyttää mahdollisimman pieneltä.



15. –

16. a) 1 b) 4 c) 5

17. a) $Q_1 \approx 17{,}8$, $Q_3 \approx 19{,}0$
b) kvartiilivälin pituus 1,2
c) 50 %

18. $Q_1 = 6$, $Q_3 = 8$

19. lineaarinen malli

20. polynominen malli

21. ei ole riippuvuuutta

22. a) 80 kJ b) 500 kJ

23. 9,770 m/s^2

24. a) $y \approx 0{,}0002x + 7{,}9104$
b) 2,5 lasta 1000 lasta kohti

25. a) A, C, D
b) A: $r \approx 0{,}87$
B: $r \approx 0{,}15$
C: $r \approx 0$
D: $r \approx 0{,}97$
c) A: suora
B: -
C: 2. asteen käyrä
D: suora

26. Kuva A: b, d, e
Kuva B: a, c, f

27. a) $0{,}5 < r < 1$
b) –

28. Regressiosuorat:
Palkansaajat:
$y = 23{,}936x - 45\,530$
Yrittäjät:
$y = 12{,}410x + 25\,013$
Mallin mukaan vuonna 1994 palkansaajia 2 198 000 kpl ja yrittäjiä 267 000 kpl.

29. Regressiosuoraa $y = 5091{,}5 + 1{,}9794x$, missä x on kahden lajin summapistemäärä ja y kokonaispistemäärä.
Korrelaatiokerroin $r \approx 0{,}619$.
Hämäläisen kahden lajin summapistemäärä on 1590 ja ennustearvo 8239 pistettä.

Hakemisto

A
aikasarja 172
alkeistapaus 91
aritmeettinen keskiarvo 43, 46

B
binomitodennäköisyys 138

F
frekvenssi 9, 10
frekvenssijakauma 9

G
geometrinen todennäköisyys 96

H
hajontaluku 43, 57
havaintoaineisto 8
havaintoyksikkö 8
histogrammi 170

J
jatkuva muuttuja 13

K
kertoma 127
kertymäfunktio 73
keskihajonta 59
keskiluku 43
keskipoikkeama 58
kokonaisaineisto 8
kombinaatio 129
komplementti 104
korrelaatiokerroin 181
kvartiili 175

L
luokkakeskus 14
luokkaväli 14

M
mediaani 30
mediaaniluokka 32
moodi 10
muuttuja 8

N
normaalijakauma 72, 73
normitettu arvo 70

O
odotusarvo 73
otantamenetelmä 9
otos 9
otoskeskihajonta 59

P
painotettu keskiarvo 50
palkkidiagrammi 169
perusjoukko 8
populaatio 8
porrasdiagrammi 27, 169
pylväsdiagrammi 168

R
regressiokäyrä 178

S
satunnaisilmiö 17
sektoridiagrammi 169
sijaintiluku 43
suhteellinen frekvenssi 9
summafrekvenssi 27, 28
summakäyrä 28
suotuisa tapaus 91

T
tapahtuma 90
tilastoaineisto 8
tilastoyksikkö 8
toistokoe 136
tuloperiaate 126
tunnusluku 43
tyyppiarvo 10

V
varianssi 58

Y
yhteenlaskusääntö 116